JMIR Diabetes

Emerging Technologies, Medical Devices, Apps, Sensors, and Informatics to Help People with Diabetes
Volume 11 (2026) ISSN 2371-4379 Editors-in-Chief: Ricardo Correa, MD, EdD; Sheyu Li, MD

Contents

Original Papers

Exploring the Needs of Health Professionals for a Type 2 Diabetes Remote Patient Monitoring Dashboard for Personalized Care: Focus Group Study (e84894)
Chiara Lansink, Nienke Beerlage-de Jong, Eclaire Hietbrink, Anouk Middelweerd, Gozewijn Laverman. 6

Bridging the Digital Divide in Diabetes Care With an AI-Powered Health Coaching Intervention: Qualitative Formative Study (e81441)
Tejossy John, Sangeetha Mohanraj, Tanjila Nawshin, Tapan Mehta, Mohanraj Thirumalai. 28

GraphRAG-Enabled Local Large Language Model for Gestational Diabetes Mellitus: Development of a Proof-of-Concept (e76454)
Edmund Evangelista, Fathima Ruba, Salman Bukhari, Amril Nazir, Ravishankar Sharma. 45

Personalized Type 1 Diabetes Management: Reinforcement Learning–Based Insulin Dosing and Glucose Forecasting (e79195)
Ernest Taku, Vibhuti Gupta, Ashutosh Singhal. 70

Integration of Continuous Glucose Monitoring With HbA1c to Improve the Detection of Prediabetes in Asian Individuals: Model Development Study (e81520)
Michelle Lee, Shihui Jin, Eveline Febriana, Maybritte Lim, Sonia Baig, Shahmir Ali, Ian Ang, Tze Loh, Ashna Nastar, Kee Chia, Alice Kong, Faidon Magkos, Alex Cook, Sue-Anne Toh. 86

App-Based Digital Therapeutics Integrating Continuous Glucose Monitoring for Glycemic Control in Type 2 Diabetes: Prospective Observational Cohort Study (e86651)
Banshi Saboo, Mudit Sabharwal, Mihir Gharia, Devina Aswal, Parth Modi, Twinkle Maheshwari, Jaymin Parikh, Simpel Saanyal, Jaikishan Agrawal. 104

Quality and Guideline Adherence of Mobile Nutrition Management Apps for Diabetes: Evaluation Study (e80890)
Sidra Jahanzeb, Nadia Davoody. 115

Exploring User Experiences With the Lift App for Emotional Well-Being Among Youth With Type 1 Diabetes: Qualitative Study (e79896)
Katie Babbott, Kristin Ginsberg, Anna Boggiss, Joanna McClintock, Chloe Hudson, Paul Hofman, Ryan Paul, Alana Cavadino, Hiran Thabrew, Craig Jefferies, Martin de Bock, Jennifer Sherr, Anna Serlachius. 133

A Serious Digital Game (SugarVita) to Support Diabetes Self-Management: Pilot Randomized Controlled Trial (e99345)
Edouard Reinders, Elma Cruts - Van de Meulengraaf, Ryan De Vries, Natal van Riel, Pieter Van Gorp, Jeanne Dieleman, Pleun Wouters - van Poppel, Harm Haak. 143

XSL•FO
RenderX

Exploring the Integration of Consumer Activity Trackers Into a Community Weight Management Intervention to Support Physical Activity in Adults at Risk for or With Type 2 Diabetes: Mixed Methods Study Using the RE-AIM Framework (e91073)
William Hodgson, Alison Kirk, Marilyn Lennon, Xanne Janssen, David Kennedy. 154

Cultural and Technological Barriers to Telehealth Adoption for Type 2 Diabetes Management Among Asian American Patients: Qualitative Case Study (e75689)
Devi States. 174

The Process of Developing an Intervention to Increase Awareness of Cardiovascular Risk for Persons With Type 2 Diabetes: Co-Creation Study (e85748)
Anna-Lena Stenlund, Karin Hellström Ängerud, Mikael Lilja, Julia Otten, Lena Jutterström. 187

Personalized Digital Health Solutions to Increasing Diabetes-Related Knowledge and Behavioral Outcomes: Results From a Randomized Controlled Trial (e87364)
Marcia Ory, Christi Esquivel, Gang Han, Samuel Towne Jr, SangNam Ahn, Matthew Smith. 203

Promoting Diabetes Self-Management Among Vietnamese Americans: Mixed Methods Pilot Study (e80177)
Anna Nguyen, Valerie Eschiti, Thanh Bui, Katherine O'Neal, Tam Nguyen, Angelina Nguyen, Hanxia Li, Michael Machiorlatti, Kathleen Dwyer.
2 2 1

Promoting Stigma-Free Language in Type 2 Diabetes Care Among Primary Care Clinicians: Randomized Controlled Pilot Trial (e95805)
Kevin Joiner, Alexandra Agapiou, Alicia Carmichael, Brandon Labbree, Donna Walter, Houda Tarraf, Medha Raju, Boluwatife Dogari, Nao Hagiwara, Sophia Lin, Richard Gonzalez. 235

Content Validation of an Electronic Health Record–Based Diabetes Self-Management Support Tool for Older Adults With Type 2 Diabetes: Qualitative Study (e83448)
Ploypun Narindrarangkura, Siroj Dejhansathit, Uzma Khan, Margaret Day, Suzanne Boren, Eduardo Simoes, Min Kim. 250

Cardiovascular-Kidney-Metabolic Syndrome: Development of an ICD-10-CM Coding Framework (e91827)
Minzhe Zhao, Junhao Zhang, Huiyun Wang, Chongjun Xu, Jiahui Li, Chenghua Li, Xuemei He, Xueying Zheng. 264

User-Centered Development of a Digital Health Service for Diabetic Foot Ulcer Risk Stratification: Usability Study (e83287)
Ulla Hellstrand Tang, Shivani Ravichandran, Stefan Candefjord, Dipu Sathyapalan, Vivek Lakshmanan. 284

The Lived Experience of Adolescents With Type 1 Diabetes Mellitus in Bahir Dar, Ethiopia: Phenomenological Study (e71179)
Yassin Hussin, Desalegn Awle, Hailemariam Mekonnen Workie, Desalegne Amare. 299

Age- and Sex-Specific Distribution of the Triglyceride-Glucose Index in a Large Chinese Population: Cross-Sectional Study (e95855)
Hanze Du, Shuaihua Song, Xiao Zhai, Junchi Liu, Yue Jiang, Yaoda Hu, Huijing He, Daowei Li, Huijuan Zhu, Shi Chen, Guangliang Shan, Hui Pan. 312

Experiences of Adults With Type 1 Diabetes Using Digital Health Technology for Diabetes Self-Care: Qualitative Study (e79704)
Divya Stephen, Jan Nilsson, Unn-Britt Johansson, Anna Nordin. 342

Social Media Discussions of Anti-Diabetic Drugs and the Popularity of GLP-1 Therapies: Content Analysis (e87590)
Chloe Stallion, Tamkeen Khan, Christian Guerrero, Stavros Tsipas, Gregory Wozniak. 355

XSL•FO
RenderX

Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) and Hospital-Based Approaches to Closed-Loop Therapy in Type 1 Diabetes: Cost-Effectiveness Analysis (e86690)
Mercia Napame, Sylvie Picard, Tony Foglia, Anne Rubenstrunk, Florence Baudoux, Francoise Giroud, Sandrine Lablanche, Sophie Borot. . . 5 6 3

Use of the GTT@home Oral Glucose Tolerance Test Kit in Gestational Diabetes Mellitus: Performance Evaluation Study (e69695)
Gareth Dunseath, Stephen Luzio, Wai Cheung, Sharon Parsons, Nicola John, Mahmoud Chokor, Michael Atkinson, Rajesh Peter. 579

Social Determinants of Health Among Chinese Americans at Risk for Diabetes in a Mobile Diabetes Prevention Trial: Cross-Sectional Baseline Analysis (e95295)
Nan Jiang, Jing Liu, Haili Song, Yanan Zhao, Huilin Li, Stella Yi, Jeannette Beasley, Lu Hu. 590

The Emotional Burden and Behavioral Impacts of Alarms and Alerts Across the Spectrum of Diabetes Technology Use: a Cross-Sectional Study (e97122)
Estelle Everett, Erin Shaw, Bryan Escobar Barrios, Beatrice Brumley, Kyrstin Lane, Tannaz Moin, Lauren Wisk. 602

Exploring the REACHOUT Mental Health Support App for Type 1 Diabetes From the Perspectives of Recipients and Providers of Peer Support: Qualitative Study (e72779)
Debbie Lam, Poonamdeep Jhajj, Diana Sherifali, Frances Chen, Tricia Tang. 616

Implementing Diabetic Eye Care in Rural Primary Care Through Teleophthalmology and Pop-Up Clinics: Mixed Methods Study (e106290)
Nisha Arya, Victoria Koltchine, Talia Gearinger, Sherianne Buehler, Reza Yousefi-Nooraie, Rajeev Ramchandran. 629

Artificial Intelligence in Diabetic Kidney Disease Research: Bibliometric Analysis From 2006 to 2024 (e72616)
Xingyuan Li, Liming Xiao, Fenghao Yang, Fang Liu. 643

Assessing Quality, Readability, and Transparency of Online and AI-Generated Information on Type 2 Diabetes Mellitus: Cross-Sectional Exploratory Study (e93822)
Sripradha Srinivasan, Ravanth Baskaran, Srinjay Mukhopadhyay, Mayank Patel. 651

Predicting Treatment Failure With Sodium-Glucose Cotransporter-2 Inhibitors in People With Type 2 Diabetes: Novel Artificial Intelligence and Machine Learning Approach (e85372)
Doyoung Kwak, Xi Tan, Yuanjie Liang, Caroline Swift, Chalak Muhammad, Xu Shi. 664

Antidiabetic Drug Associations With Heart Failure Outcomes: Real-World Evidence Study Using Electronic Health Records (e85083)
Elzbieta Jodlowska-Siewert, Yunhui Chen, Sinian Zhang, Jia Li, Robert Dellavalle, Rui Zhang, Jue Hou. 678

Cardiorespiratory Markers of Type 2 Diabetes: Machine Learning–Based Analysis (e82084)
Flavia Oliveira, Sandro Cavalcanti, Michael Khoo. 691

Predictors of Glycemic Response to Sulfonylurea Therapy in Type 2 Diabetes Over 12 Months: Comparative Analysis of Linear Regression and Machine Learning Models (e82635)
Shilpa Garg, Robert Kitchen, Ramneek Gupta, Emanuele Trucco, Ewan Pearson. 711

Privacy-Preserving Collaborative Diabetes Prediction in Heterogeneous Health Care Systems: Algorithm Development and Validation of a Secure Federated Ensemble Framework (e79166)
Md Hasan, Juan Li. 726

XSL•FO
RenderX

Research Letters

Continuous Ketone Monitoring: Data From a Randomized Controlled Trial (e85548)
Simon Kjær, Lukas Ridder, Mads Svart, Nikolaj Rittig, Lise Bjerg, Birgitte Sandfeld-Paulsen, Henrik Thomsen. 100

Toward Robust AI-Assisted Dietary Assessment for Diabetes Self-Management: Quantifying and Decomposing Large Language Model Prediction Variability From Meal Images (e102715)
Zhaohua Wang, Daniel Lane, Kayo Waki. 745

Viewpoints

Harnessing Digital Innovation for Diabetes Care: Insights From the Action4Diabetes–CorrelAid Data4Good Collaboration (e89357)
Sze Ng, Michael Aydinbas, Tyla Martin, Darren Toh, Steven James. 320

T1D REACHOUT–A Mobile App to Deliver Peer-Led Mental Health Support to Adults Living With Type 1 Diabetes: Co-Design and Development Process (e71733)
Sadie Lee, Baray Sidhu, Parteek Johal, Ayman Azhar, Jonath Sujan, Matthias Görges, Tricia Tang. 325

Reviews

Community-Acquired Pneumonia in Patients With Diabetes: Narrative Review (e82215)
Yun Xie, Ao Zhang, Ying Wang, Ruilan Wang. 366

The Evolving Landscape of Continuous Glucose Monitoring Metrics in Type 1 Diabetes: Narrative Literature Review (e92455)
Eric Williams Jr, Casey Rand, Alessandra Ayers, Cindy Ho, Ashley Dunova, David Klonoff, Juan Espinoza. 382

Wearable Technologies for Diabetic Foot Ulcer Monitoring and Risk Prediction: Systematic Review (e84463)
Houriyeh Ehtemam, Simin Salehinejad, Dan Robbins, Alireza Sanaei, Hassan Shirvani, Shabnam Sadeghi-Esfahlani. 398

AI-Based Models for Diabetic Foot Ulcer Assessment: Scoping Review (e77925)
Muhamad Zulfiqar, Saldy Yusuf, Muhammad Taming, Herlina Burhan. 418

Digital Twin Applications in Diabetes Management: Scoping Review (e83059)
Fatemeh Sarani Rad, Maryam Jafarpour, Ehsan Bitaraf, Katayoon Khaleghdadi, Juan Li. 437

Training Community Health Workers for Diabetes Management in Low- and Middle-Income Countries: Systematic Review (e84508)
Anirudh Gudlavalleti, Sureshkumar Kamalakannan, Sophie Vanbelle (S), Venkata Gudlavalleti, Nicolaas Schaper, Giridhara Babu, Onno Schayck. 458

Obstacles and Enablers Related to Gestational Diabetes Self-Management: Systematic Review Using the Socio-Eecological Model (e86767)
Bronya Luk, Polly Ma, Rosenna Chung, Wendy Zhang, William Lam. 474

Digital Health Solutions for Type 2 Diabetes and Prediabetes: Systematic Review of Engagement Barriers, Facilitators, and Outcomes (e80582)
Ayesha Thanthrige, Nilmini Wickramasinghe. 491

XSL•FO
RenderX

Effectiveness, Reach, Uptake, and Feasibility of Digital Health Interventions for Culturally and Linguistically Diverse Populations Living With Prediabetes Across the Lifespan: Systematic Review and Meta-Analysis (e70912)
Lisa Whitehead, Min Zhang, Wai Kwok, Diana Arabiat, Amanda Towell Barnard. 508

Continuous Glucose Monitoring–Derived Metrics and Cardiovascular Risk Among People With Diabetes: Systematic Scoping Review (e89374)
Helene Thomsen, Benjamin Lebiecka-Johansen, Ole Nørgaard, Tue Helms Andersen, Signe Andersen, Guy Fagherazzi, Adam Hulman, Anders Isaksen. 526

Tutorial

A Quantitative Framework for Evaluating the Performance of Algorithm-Directed Whole-Population Remote Patient Monitoring: Tutorial for Type 1 Diabetes Care (e72676)
Jamie Kurtzig, Ananta Addala, Franziska Bishop, Paul Dupenloup, Johannes Ferstad, Ramesh Johari, David Maahs, Priya Prahalad, Dessi Zaharieva, David Scheinker. 545

XSL•FO
RenderX

Exploring the Needs of Health Professionals for a Type 2 Diabetes Remote Patient Monitoring Dashboard for Personalized Care: Focus Group Study

Chiara Lansink[1,2], MSc; Nienke Beerlage-de Jong[3], PhD; Eclaire Hietbrink[1,2], PhD; Anouk Middelweerd[1], PhD; Gozewijn Dirk Laverman[1,2], MD, PhD

[1]Department of Biomedical Signals and Systems, Univerisity of Twente, Enschede, The Netherlands
[2]Department of Internal Medicine/Nephrology, Ziekenhuisgroep Twente, Almelo, The Netherlands
[3]Section of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands

Corresponding Author:
Chiara Lansink, MSc
Department of Biomedical Signals and Systems, Univerisity of Twente, Enschede, The Netherlands

Abstract

Background: Effective management of type 2 diabetes mellitus (T2DM) requires monitoring clinical parameters like blood glucose and medication, alongside lifestyle factors such as diet and physical activity. Decision support tools, including dashboards and shared decision-making tools, help with medication adjustments, glucose monitoring, and lifestyle. However, systems rarely integrate home-monitored lifestyle data with personalized guidance and rarely facilitate collaborative goal setting for behavior change. As a result, health care professionals (HCPs) are limited in their ability to support patients' medical and lifestyle management. Blended care, combining in-person consultations with digital monitoring of patient data, can help bridge this gap by providing structured information and data-driven insights to support diabetes management.

Objective: The study aims to identify HCPs' requirements for a remote monitoring dashboard for people with diabetes that integrates clinical and home-monitored lifestyle data, supporting personalized, patient-centered care and collaborative goal setting in blended T2DM management.

Methods: A qualitative study was conducted using 2 interactive focus group sessions with HCPs involved in the treatment of T2DM. Focus group participants shared experiences, identified practical needs, and collaboratively defined requirements for a dashboard to support personalized diabetes management. Transcripts were coded to identify recurring themes and ideas, which were then consolidated into distinct requirements. Requirements were labeled with an identification code (ID) and categorized in accordance with the FICS framework, distinguishing 4 types of design requirements: functions and events (F), interaction and usability (I), content and structure (C), and style and aesthetics (S). Prioritization of requirements was performed using the must/should/could/will not have method.

Results: In total, 9 HCPs participated in 2 focus groups, each lasting approximately 1.5 hours. A total of 50 requirements for a T2DM dashboard were identified. Of these, 31 (62.0%) were functions and events (F), 9 (18.0%) related to interaction and usability (I), 7 (14.0%) concerned content and structure (C), and 3 (6.0%) pertained to style and aesthetics (S). The participants expressed the need for a dashboard that incorporates data-driven lifestyle (eg, physical activity and nutrition) with visual trend analysis and integration of psychosocial aspects. They also emphasized the importance of visualizing how nutrition, physical activity, and medication interact to influence glucose values. In addition, participants expressed the wish for a home screen that provides a quick overview, with the option to click through to more detailed views (eg, per day, week, or month).

Conclusions: The findings demonstrate a demand among HCPs for an integrated dashboard that combines clinical, lifestyle data, and psychological data to support personalized, patient-centered care. By linking lifestyle behaviors with glucose outcomes, such a tool could support collaborative goal setting, strengthen blended care pathways, and promote data-driven diabetes management. The findings provide guidance for the design of digital health interventions tailored to the needs of HCPs.

(JMIR Diabetes 2026;11:e84894) doi:10.2196/84894

KEYWORDS

type 2 diabetes mellitus; clinical decision support system; health information technology; health personnel; patient-centered care; digital health; blended care; dashboard

XSL•FO
RenderX

Introduction

Diabetes is one of the most common chronic diseases worldwide [1]. In 2022, approximately 1.2 million people in the Netherlands had diabetes, with the majority being diagnosed with type 2 diabetes mellitus (T2DM) [2,3], and this number is expected to rise to 1.5 million by 2040 [2]. This places a growing burden on health care systems and increases the need for effective prevention and management strategies [4]. T2DM is strongly influenced by lifestyle factors, with unhealthy behaviors playing a major role in its development and progress [5]. Maintaining a healthy lifestyle, including regular physical activity and balanced nutrition [6-8], as well as stress management and adequate sleep [9,10], is crucial for glycemic control and overall health outcomes. However, despite the well-known benefits of a healthy lifestyle, recent studies showed that 49% of the Dutch population did not meet the recommended physical activity guidelines [11,12], and similar patterns are observed in persons with diabetes. Adherence to both nutritional and physical activity recommendations is low, with only a small proportion of persons with diabetes meeting the guidelines [13-16].

Addressing lifestyle as part of treatment is emphasized in the diabetes guidelines [17,18]. Health care professionals (HCPs) are expected to support persons with diabetes in adopting healthier behaviors by assessing lifestyle factors, offering guidance, and helping individuals prioritize achievable changes [19]. In this context, HCPs serve as the critical link in translating lifestyle recommendations into practical advice that persons with diabetes can apply in daily life. However, routine practice consultations often focus primarily on glucose levels and evaluation of medication effects on parameters such as blood pressure and cholesterol [20,21], with limited attention to lifestyle counseling due to time constraints [20,22].

To help address these challenges, digital tools have been increasingly explored as supportive tools in diabetes care. Reviews highlight the potential of mobile health apps, continuous glucose sensors, and activity trackers to support lifestyle changes and diabetes self-management [23-27]. In addition to monitoring lifestyle, decision-support systems and remote persons with diabetes monitoring have been shown to help reduce pressure on health care services by decreasing outpatient visits and consultation times and reducing hospital admissions [28,29]. At the same time, HCPs consistently highlight the value of face-to-face contact [30-32], suggesting that digital solutions are most useful when embedded within blended care models. Such models, in which digital tools are integrated into in-person consultations, are supported by evidence, indicating that digital innovations are more effective when complemented by a human component [33-37].

One particularly promising digital tool to be used in such blended care is a dashboard that facilitates shared decision-making by presenting relevant health data in a structured and visually accessible format. Shared decision-making allows HCPs and persons with diabetes to jointly tailor care plans to individual preferences and values [38]. To the best of our knowledge, no existing dashboard integrates lifestyle behaviors such as diet and physical activity with clinical metrics like continuous glucose monitoring into personalized lifestyle advice in a data-driven way. This hinders the delivery of data-driven care that goes beyond glucose regulation alone. The development of a dashboard that combines these data streams into tailored recommendations could therefore provide a practical blended care solution. It could strengthen shared decision-making and support personalized diabetes management in routine care.

Therefore, the objective of this study was to identify and determine the requirements of HCPs for a remote patient monitoring dashboard that visualizes health data (eg, glucose regulation, physical activity, and nutrition) to support personalized blended care and shared decision-making. Through focus groups, we investigated which information HCPs consider most relevant, how they prefer it to be presented, and what additional data would make the dashboard most clinically useful. The ultimate goal is to enable HCPs and persons with diabetes to collaboratively set personalized lifestyle goals and make informed treatment decisions, thereby facilitating more individualized care within blended care settings.

Methods

Design

This study used a stepwise approach, collecting data through 2 interactive focus groups. These sessions facilitated both the sharing of experiences and collaborative identification of practical needs for the dashboard, making this approach well-suited for designing user-centered digital health interventions. The focus groups were conducted following Raats' guidelines for health care settings [39].

Ethical Considerations

The study was reviewed and approved by the ethics committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente (reference: 240227). Written informed consent was obtained from all participants prior to the initiation of the focus groups, and they were fully informed about the purpose of the study. Participants' privacy and confidentiality were strictly maintained throughout the research. No compensation was provided for participants. At the beginning of each session, participants were reminded that it would be recorded.

Participants and Recruitment

We used purposive sampling to recruit HCPs involved in T2DM care across secondary and primary care in the Twente region (Netherlands). In total, 27 HCPs were invited across both sessions. In the first focus group, 8 HCPs from secondary care, together with a lifestyle coach, were invited to participate in an initial focus group, aimed at exploring needs and challenges in secondary care. All of them were welcome to participate in both sessions. In total, 16 HCPs were approached for the second focus group, including professionals from the 2 regional hospitals and primary care providers. This study had an exploratory qualitative design. The aim was to generate new insights rather than to reach statistical generalizability.

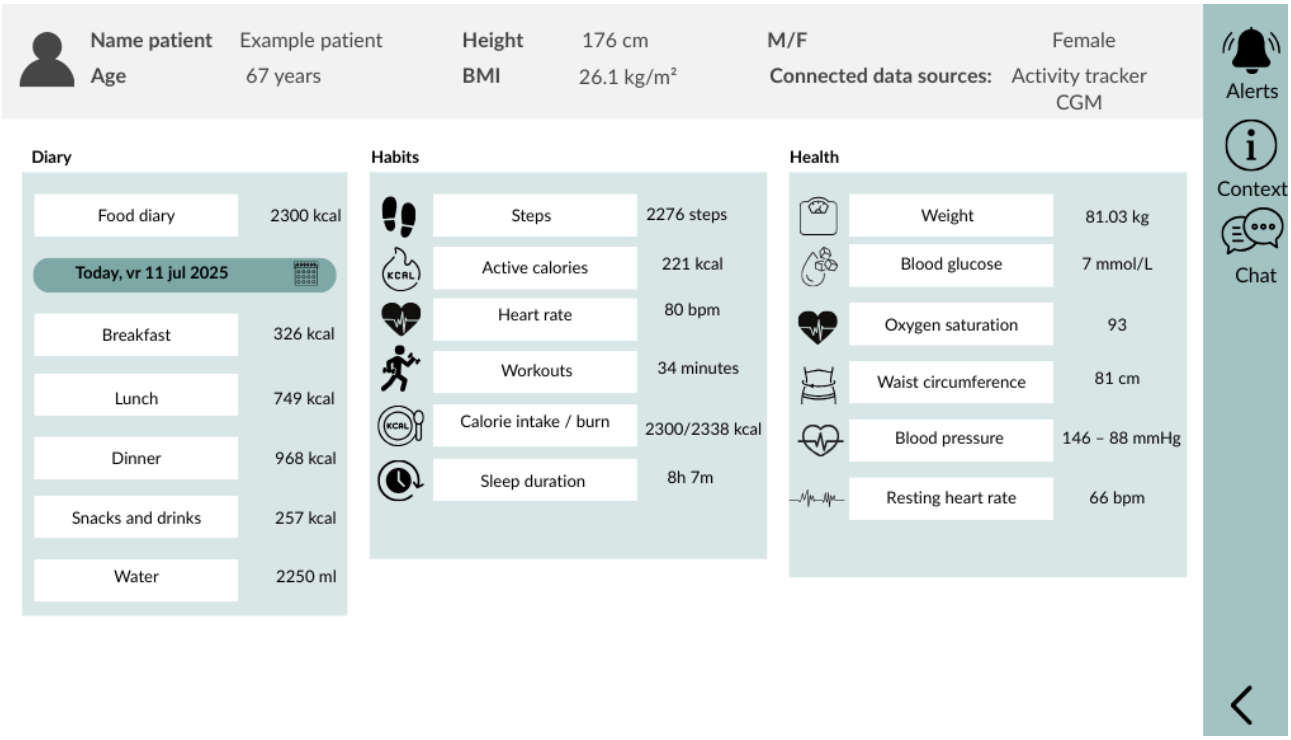
XSL•FO
RenderX

Consequently, a smaller yet information-rich sample of diabetes nurses was considered methodologically appropriate [40].

Data Collection

Two 1.5-hour focus groups were conducted following established guidelines for health care–based group discussions and using a structured requirements development approach. Each session began with introductions and followed a predefined protocol covering current monitoring practices, desired data visualization, and workflow integration. In the first session, participants mapped the current and ideal care pathways using a hypothetical referral case and identified corresponding dashboard needs. They then individually listed and ranked their requirements, which were jointly reviewed and prioritized using the must/should/could/will not have (MoSCoW) approach [41]. Based on requirements obtained during the first session, a mock-up dashboard was created (Figure 1).

Figure 1. Visualization of a mock-up dashboard that was shown to the health care professionals during the second focus group.



In the second session, participants designed an initial dashboard by selecting and ranking essential parameters, compared these with previously identified requirements, and provided feedback on an interactive mock-up. The session also explored challenges and opportunities for transmural (primary-secondary) collaboration and the potential role of the dashboard in blended care. The researchers closely monitored group dynamics and ensured a safe environment in which participants could speak up and discuss differences in opinion.

Data Analysis

Audio recordings were transcribed verbatim using Amberscript and corrected manually. Session summaries were member-checked by participants. We applied a 9-step requirements development method [42]: familiarization, selection of relevant quotes, attribute assignment, grouping and refinement, translation into preliminary requirements, independent review, and consensus building. Relevant quotes were identified using Braun's selection criteria [43]. An independent researcher (NB-J) provided a second opinion review to strengthen the reliability of the analysis.

To organize the final set of requirements, we used the FICS structure to present results from a design-oriented perspective [44]. This framework distinguishes 4 types of design requirements: functions and events (F), referring to the core features needed for the dashboard to operate; interaction and usability (I), relating to user experience and accessibility; content and structure (C), focusing on how information is organized and presented; and style and aesthetics (S), addressing visual design and appeal. Prioritization of the resulting requirements followed the MoSCoW categories; in the first focus group, these were assigned directly by participants, and in the second focus group, they were interpreted and categorized by the researcher (CL).

Results

Characteristics of the Participants

In total, 11 participants agreed to take part in the focus groups. Due to illness (n=2) and personal circumstances (n=1), eventually, 9 participants attended the sessions, one of whom joined both sessions (n=10). Each session included 5 participants. The group consisted of 6 female and 3 male participants, including 3 diabetes nurses, 1 diabetes specialist nurse, 3 internists, and 2 general practitioner practice assistants. Of these, 2 participants were from primary care and 7 from secondary care. The 3 participating internists each had a different subspecialty: 1 in endocrinology, 1 in nephrology, and 1 in vascular medicine. An overview of the participants' characteristics is provided in Table 1.

XSL•FO
RenderX

Table . Unique focus group participants (N=9)[a].

Characteristics		Participants, n (%)
Sex		
	Female	6 (66.7)
	Male	3 (33.3)
Profession		
	Diabetes nurse	3 (33.3)
	Diabetes nurse practitioner	1 (11.1)
	Internist	3 (33.3)
	General practitioner practice assistant	2 (22.2)
Type of care		
	Primary care	2 (22.2)
	Secondary care	7 (77.7)
Years of experience in diabetes care		
	0 - 10	5 (55.6)
	11 - 20	1 (11.1)
	20+	3 (33.3)

[a]In total, 1 participant joined both focus groups, resulting in a total of 9 unique participants.

Overview of the Total Requirements

A total of 50 requirements were identified. These requirements were categorized into 4 main groups (see also Table 2): 31 (62.0%) functions and events (F) requirements, 9 (18.0%) interaction and usability (I) requirements, 7 (14.0%) content and structure (C) requirements, and 3 (6.0%) style and aesthetics (S) requirements.

Table . The number of formulated requirements as a result of the focus groups per FICS[a] category for overarching, dietary intake, physical activity, glucose values, social domain, and additional parameters.

	Overall system, n (%)	Dietary intake, n (%)	Physical activity, n (%)	Glucose values, n (%)	Psychosocial aspects, n (%)	Additional parameters, n (%)	Total, n (%)
Functions and events	10 (20.0)	3 (6.0)	3 (6.0)	10 (20.0)	3 (6.0)	2 (4.0)	31 (62.0)
Interaction and usability	5 (10.0)	0 (0.0)	1 (2.0)	3 (6.0)	0 (0.0)	0 (0.0)	9 (18.0)
Content and structure	1 (2.0)	2 (4.0)	0 (0.0)	0 (0.0)	1 (2.0)	3 (6.0)	7 (14.0)
Style and aesthetics	2 (4.0)	0 (0.0)	0 (0.0)	1 (2.0)	0 (0.0)	0 (0.0)	3 (6.0)
Total	18 (36.0)	5 (10.0)	4 (8.0)	14 (28.0)	4 (8.0)	5 (10.0)	50 (100.0)

[a]FICS: Functions and events, interaction and usability, content and structure, and style aesthetics.

The requirements covered overall system requirements, including accessibility of the dashboard as well as requirements specifying the design of the dashboard. They also included specific requirements for the monitoring of dietary intake, physical activity, glucose values, psychosocial aspects of living with T2DM, and additional parameters (eg, laboratory results and sleep).

A total of 37 (74.0%) requirements were classified as “must have” or “should have,” indicating their essential role in the dashboard’s core functionality. The remaining 13 (26.0%) requirements fell into the “could have” or “will not have” categories, meaning that they were not essential for basic functionality or were excluded from this phase to avoid overcomplicating the design. An overview of all requirements and their classification can be found in Tables 3-8.

XSL•FO
RenderX

Table . Requirements regarding the overall system[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F1	The dashboard is accessible to all HCPs[b] involved in diabetes management (eg, diabetes nurses, primary care physicians, and specialists).	Diabetes management often involves multiple HCPs.	✓			
F2	The dashboard supports integration with multiple health apps.	This is considered important for long-term sustainability and broader adoption of the dashboard.				✓
F3	There is a home screen.	A home screen provides fast information to the HCPs.	✓			
F4	The home screen includes data about glucose values, nutritional intake, physical activity, and medication use.	This information provides a clear century point for the consultations.	✓			
F5	The home screen includes a clickable structure that allows HCPs to easily navigate to more detailed data.	A clickable structure keeps the balance between overview and depth.		✓		
F6	Persons with diabetes can be monitored over a specific period.	Allows HCPs to review specific time frames to evaluate treatment effectiveness.	✓			
F7	Integrated chat feature allows persons with diabetes to easily contact health care providers and vice versa.	Improves communication and enables timely responses between consultations.			✓	
F8	The chat feature automatically sends notifications to HCPs.	Ensures messages are seen promptly.			✓	
F9	Automated messages are sent to persons with diabetes based on their physical activity, nutrition, and glucose levels using artificial intelligence.	This represents potential future improvement for automated personalized diabetes care.				✓

XSL•FO
RenderX

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
F10	Data can be anonymized for use in training.	Enables the use of real cases for education and system testing while protecting privacy and complying with ethical standards.			✓	
Interaction and usability						
I1	The dashboard can integrate into existing software systems.	This supports smooth adoption in daily practice.	✓			
I2	The dashboard can connect with multiple electronic health record systems.	This ensures interoperability across primary care and secondary care HCPs.	✓			
I3	The dashboard provides easily understandable information at a glance.	Quick and clear insights support engagement without being overwhelmed by data.	✓			
I4	The dashboard enables HCPs to select which parameters are visible, allowing for customization based on agreements with the person with diabetes.	Tailored dashboards prevent information overload.		✓		
I5	Navigation within the dashboard follows a logical and consistent structure.	Navigating in the dashboard should align with clinical workflows.	✓			
Content and structure						
C1	The health record of a person with diabetes is accessible for different HCPs involved, meaning that the health record is accessible across care domains and pathways.	Promotes interdisciplinary collaboration.		✓		
Style and aesthetics						
S1	The home screen can be personalized by HCP, based on their role, specialization, or preferences.	Not all HCPs need the same information; for example, an internist has different priorities than a nurse. Personalization increases the relevance of the information displayed.		✓		

XSL•FO
RenderX

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
S2	The dashboard includes an overview of historic data of persons with diabetes in an overview per day, per week, and per month.	To see the progression over time.		✓		

[a]Requirements are categorized as functions and events (F), interaction and usability (I), content and structure (C), and style and aesthetics (S). Each has an identification code (ID) in the first column, matching references in the main text.

[b]HCP: health care professional.

XSL•FO
RenderX

Table . Requirements regarding dietary intake[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F11	The dashboard enables the display of an overview of the consumed carbohydrates, amount of fat, and proteins.	This overview allows HCPs[b] to quickly assess general nutritional intake.	✓			
F12	Data output (see F11) is stored per day and presented for breakfast, lunch, dinner, in-between meals, and total carbohydrate intake.	To gain insight into carbohydrate distribution over the day.		✓		
F13	The dashboard provides HCPs with basic nutritional information and practical guidelines, such as the percentage of dietary reference intake for relevant nutrients (see F11).	To enable HCPs to confidently address general nutrition questions.	✓			
Content and structure						
C2	Data output is presented in grams per day.	For each product, the amount of, for example, carbohydrates in grams per day should be calculated, based on nutritional values per 100 grams of the food item as noted in the Dutch Food Composition Table.	✓			
C3	Nutritional data are displayed in detail, breaking it down into carbohydrates, proteins, fat, fluid intake, and alcohol consumption.	Detailed breakdowns provide deeper insights into specific nutritional behaviors.	✓			

[a]Requirements are categorized as functions and events (F) and content and structure (C). Each has an identification code (ID) in the first column, matching references in the main text.

[b]HCP: health care professional.

XSL•FO
RenderX

Table . Requirements regarding physical activity[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F14	The dashboard can display MVPA[b] levels and step count.	MVPA provides insight into activity intensity, which is essential for evaluating whether persons with diabetes meet physical activity guidelines.	✓			
F15	The dashboard displays physical activity through step counts and MVPA for daily, weekly, monthly, and yearly periods.	Insight into long-term trends of physical activity data and variations in activity levels over time allows HCP[c] for more personalized monitoring.	✓			
F16	The dashboard displays all types of physical activity, including sports do not capture through step count.	Provides a more complete picture of the activity level.		✓		
Interaction and usability						
I6	Persons with diabetes can manually input physical activities that are not automatically tracked by an activity device.	Ensures untracked but relevant activities are included.		✓		

[a]Requirements are categorized as functions and events (F) and interaction and usability (I). Each has an identification code (ID) in the first column, matching references in the main text.

[b]MVPA: moderate to vigorous physical activity.

[c]HCP: health care professional.

XSL•FO
RenderX

Table . Requirements for glucose values[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F17	The dashboards display the percentage of time in range, time below range, and time above range.	This allows HCPs[b] to quickly assess glucose levels.	✓			
F18	The dashboard displays the glucose values throughout the day and the progress over time.	This insight helps HCPs to easily track glucose control of people with T2DM[c].	✓			
F19	The dashboard displays glucose variability during the day.	This information helps HCPs identify specific times or behaviors that may contribute to instability.	✓			
F20	The dashboard can support artificial intelligence–based adjustments of medication based on glucose sensor data.	This represents a potential future improvement for automated personalized diabetes care.				✓
F21	The dashboard can synchronize with various glucose monitoring devices.	This makes it usable for a broader target population.			✓	
F22	Blood glucose value thresholds are adjustable.	The thresholds can differ per person with diabetes.	✓			
F23	The dashboard records and displays how often a person with diabetes scans their glucose sensor.	This information helps HCPs identify potential issues such as excessive scanning.			✓	
F24	Notifications with general advice are sent automatically from the dashboard to the person with diabetes when glucose values are out of range.	Notifications with advice help persons with diabetes quickly manage abnormal glucose levels.			✓	
F25	Alerts are only sent to HCPs in specific, well-defined cases.	HCPs should not be overwhelmed with notifications.	✓			
F26	Alerts forwarded to HCPs are filtered according to predefined clinical protocols.	Unfiltered alert risk will overwhelm HCPs with unnecessary notifications, which may reduce efficiency and contribute to alarm fatigue.		✓		
Interaction and usability						

XSL•FO
RenderX

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
I7	The dashboard provides a clickable option on the home screen to access a detailed view of glucose data.	Offering easy access to a more detailed view of glucose data (F17 and F18).		✓		
I8	The dashboard supports the wireless transfer of glucose data from the devices of persons with diabetes.	Wireless transfer ensures that the HCPs have access to glucose data.	✓			
I9	Persons with diabetes are considered primarily responsible for acting based on their glucose data.	HCPs cannot continuously present to monitor or respond to real-time fluctuations.	✓			
Style and aesthetics						
S3	The detailed glucose overview clearly displays how activity, nutrition, and medication influence blood glucose levels in an easy-to-understand way.	Presenting glucose values in relation to this provides deeper insight into glucose regulation.		✓		

[a]Requirements are categorized as functions and events (F), interaction and usability (I), and style and aesthetics (S). Each has an identification code (ID) in the first column, matching references in the main text.

[b]HCP: health care professional.

[c]T2DM: type 2 diabetes mellitus.

XSL•FO
RenderX

Table . Requirements for psychosocial aspects[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F27	The dashboard includes information on relevant social support within the region.	HCPs[b] lack visibility in available social resources. Integration into the dashboard makes this insightful and increases the likelihood of being referred to it.			✓	
F28	The dashboard includes a visualization of the stress levels of persons with diabetes.	Visually displaying stress can help HCPs get an overview of the impact of psychological factors on glucose regulation.		✓		
F29	Persons with diabetes can add notes to their measurements for health care providers.	Notes provide context for the measurements.			✓	
Content and structure						
C4	The dashboard displays the positive health spider web.	This ensures that HCPs can easily interpret trends in well-being.		✓		

[a]Requirements are categorized as functions and events (F) and content and structure (C). Each has an identification code (ID) in the first column, matching references in the main text.

[b]HCP: health care professional.

XSL•FO
RenderX

Table . Overview of the requirements with justification and priority levels[a].

ID	Requirement	Justification	Must have	Should have	Could have	Will not have
Functions and events						
F30	The dashboard displays blood pressure and weight.	Displaying blood pressure and weight helps HCPs[b] to get a broader overview		✓		
F31	The dashboard includes insights into sleep patterns, including total sleep duration and sleep quality.	Insights into sleep-related data to support its impact on glucose regulation		✓		
Content and structure						
C5	The dashboard provides access to medication data, including insulin use.	Insight into medication data including insulin is important to provide actual care, without the need to use HIX[c] or HIS[d]	✓			
C6	The dashboard integrates results of clinical laboratory tests.	Having access to relevant laboratory results directly within the dashboard enables HCPs to get a complete overview without needing to switch to other systems like HIX or HIS			✓	
C7	The dashboard shows the progress of weight over time.	To give insight into the weight changes			✓	

[a]Requirements are categorized as functions and events (F) and content and structure (C). Each has an identification code (ID) in the first column, matching references in the main text.

[b]HCP: health care professional.

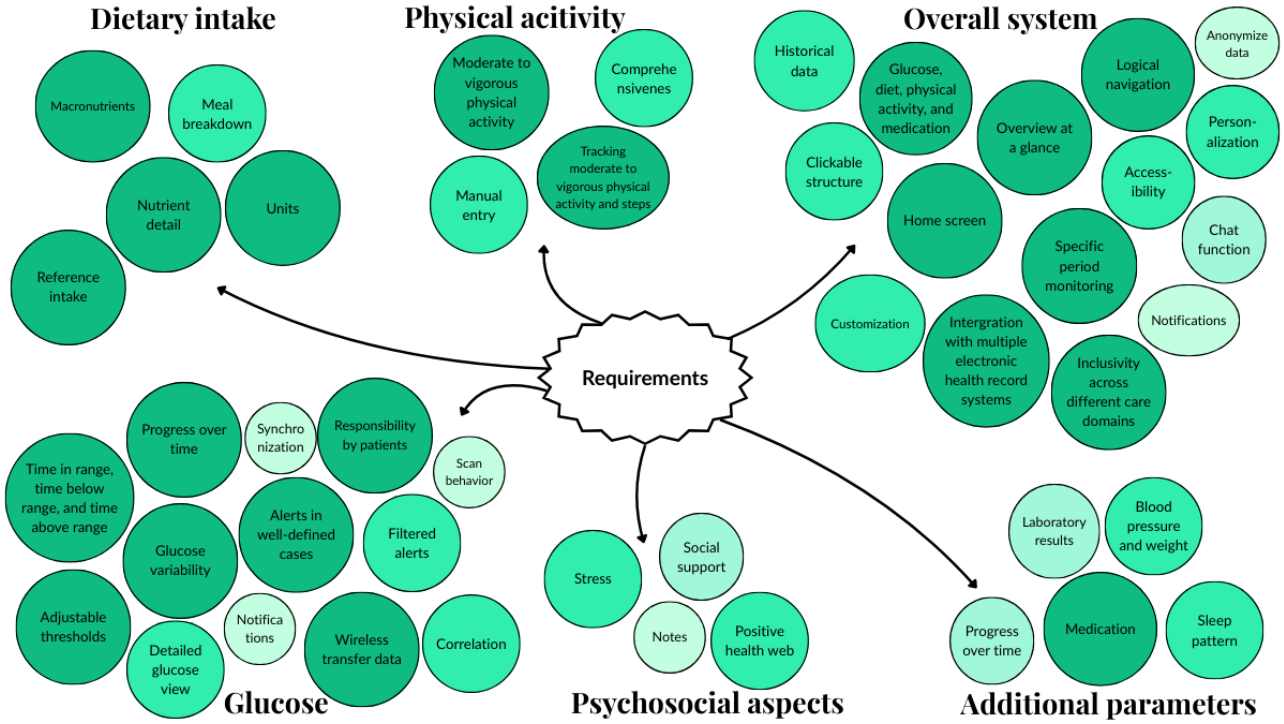
[c]HIX: Healthcare Information Exchange.

[d]HIS: Huisarts Informatie Systeem (General Practitioner Information System).

Figure 2 provides an overview of all identified requirements for the dashboard. The color and size of each element indicate its priority according to the MoSCoW classification [41]; the darker and larger the element, the higher its importance.

XSL•FO
RenderX

Figure 2. Overview of dashboard requirements.



Overall System Requirements

A set of 18 overarching requirements relates to the general functionality, usability, and accessibility of the dashboard, as shown in Table 3. These requirements ensure that the system can be used efficiently by HCPs across different disciplines and settings. In addition, participants emphasized the need to embed the dashboard within existing care pathways. Rather than serving as a stand-alone tool, the dashboard should support blended and transmural care by fostering collaboration across care levels. Participants noted that current diabetes care often lacks integrated digital solutions to monitor and address lifestyle-related factors, underscoring the importance of tools that facilitate communication and shared decision-making.

To support this, the dashboard should be accessible to all HCPs involved in diabetes management, including diabetes nurses, primary care physicians, and specialists (F1). Interoperability was considered essential to enable joint monitoring of data of persons with diabetes, establish personalized goals with the person with diabetes, and ensure seamless transitions between primary and secondary care. To achieve this and ensure integration with existing care pathways, the dashboard should be able to connect to existing software systems (I1) and multiple electronic health record systems (I2), as different care levels typically use different platforms. For sustainable implementation, it is also important that the dashboard can connect to additional health apps commonly used by persons with diabetes (F2). A central component is the home screen (F3), which should provide an accessible and simplified overview of key parameters relevant for the individual HCP (S1). Participants identified glucose values, nutritional intake, physical activity, and medication use as the most important parameters for this overview (F4), which will be explained in more detail later. To support quick interpretation and minimize unnecessary complexity, information should be presented in a concise and straightforward manner (I3): “It shouldn’t be too much, of course, otherwise you won’t get a quick overview” (Participant 1, focus group 2).

From this overview, HCPs should be able to easily navigate to more detailed data if needed. This clickable structure (F5) supports quick access to relevant measurements without overwhelming the user:

> *You should be able to click through it quickly. And if you really want to zoom in, like on a particular day, to see what exactly happened, maybe that requires a few more clicks. But even then, you should instantly get digestible averages and values to discuss with patients.* [Participant 1, focus group 1]

To prevent information overload and enhance relevance, the home screen should be customizable. HCPs should be able to select which parameters are displayed, based on personal goals (eg, exercise more) and/or circumstances (eg, comorbidities) of each person with diabetes (I4). In line with this, the dashboard should allow users to monitor the progress of persons with diabetes over a specific time period (F6). This prevents overload from entering data for persons with diabetes and from reviewing extensive information for HCPs.

> *Certain key point for these four months. So not everything necessarily needs to be tracked all the time, some aspects should always be monitored, while others will be the primary focus for the upcoming four months.* [Participant 5, focus group 2]

Additionally, historical data should be easy to navigate and visually structured, allowing HCPs to switch between daily, weekly, or monthly views depending on the context (S2). Frequently used functionalities should be directly accessible,

XSL•FO
RenderX

and overall navigation should remain intuitive to support clinical workflows without adding burden (I5).

Participants also mentioned the importance of communication between persons with diabetes and HCPs. They suggested that a built-in chat function (F7) could serve as a replacement for the email communication currently used. However, one HCP mentioned that without automatic notifications, they would have to continuously proactively check the chat, which is not feasible (F8): “But then as a healthcare provider, you really need to receive an email or something, otherwise you have to check everything manually, which just doesn’t work” (Participant 2, focus group 2).

A potential future feature for the dashboard was mentioned to send persons with diabetes personalized messages based on their physical activity, nutrition, and glucose data using artificial intelligence (AI; F9). Finally, participants would appreciate having an option to anonymize data for training or research purposes available (F10).

Requirements Regarding Monitoring Dietary Intake

Five requirements concern the monitoring of nutritional data, as shown in Table 4. These include the ability to provide an overview of consumed carbohydrates, amounts of fat, and proteins (F11), representing the core nutritional metrics needed for basic monitoring. There is also a need to present this information per meal moment (breakfast, lunch, dinner, and in-between meals; F12) and as daily totals in grams (C2). The focus groups further indicated that HCPs would appreciate a more comprehensive and detailed breakdown of nutritional data beyond these core metrics. Specifically, they suggested including additional aspects such as fluid consumption and alcohol use, alongside the existing breakdown of macronutrients (C3). This level of detail is considered necessary to support more targeted dietary counseling.

Another key theme that emerged from the focus groups was the lack of confidence among HCPs in providing nutritional advice, particularly when persons with diabetes ask specific or complex questions. As one participant reflected:

> *I sometimes find it difficult to give good advice in this area, because I don’t have a background in dietetics. And sometimes people have very specific questions, and I think, oh, yes ... but I don’t really have the exact answer either. ... I’m still trying to figure out what I can do with it [the nutritional data].* [Participant 1, focus group 1]

and

> *As doctors, at least that’s how I experience it, I really feel that I lack sufficient knowledge when it comes to nutrition. I just don’t think I know enough to give proper advice on it.* [Participant 4, focus group 1]

In response to these concerns, the system should provide easily accessible information about nutritional guidelines (F13), such as the percentage of dietary reference intake for relevant nutrients (see F11). However, one participant also questioned whether it is truly the role of HCPs in secondary or primary care to provide such detailed nutritional information, considering that nutrition is a specialized field. The participant suggested that HCPs might only need to have a certain basic level of knowledge. Such remarks do not directly translate into design requirements for the dashboard but rather highlight important implementation considerations regarding professional roles and training needs.

Requirements for Physical Activity

Four requirements relate to the presentation and interpretability of physical activity data within the dashboard, as shown in Table 5. HCPs in the focus groups expressed a preference for metrics that are easy to interpret and actionable in clinical practice. They specifically preferred using step counts and minutes of moderate to vigorous physical activity rather than calorie-burn data (F14): “Minutes of exercise, moments, frequency during the week, and minutes” (Participant 3, focus group 1).

It was recommended that both step counts and moderate to vigorous physical activity be displayed across multiple time frames (daily, weekly, monthly, and yearly; F15), enabling HCPs to track progress over time. In addition, HCPs emphasized the importance of being able to view other types of physical activity, such as swimming or cycling (F16). To ensure a complete picture of physical activity, HCPs also recommended including an option for persons with diabetes to manually add activities that may not be captured by standard trackers (I6).

Requirements for Glucose Values

A total of 14 requirements concern the monitoring of glucose values, as shown in Table 6. To provide personalized advice on glucose levels, it is essential for HCPs to have access to a recent (ie, approximately the past 2 - 4 weeks) and clearly presented overview of glucose data, reflecting the most current measurements of persons with diabetes (I7). Therefore, glucose values must be prominently displayed on the home screen of the dashboard, including key indicators such as the percentage of time in range, time below range, and time above range (F17), as well as trends over time (F18).

The clickable detailed view should offer deeper insight into the glycemic patterns of persons with diabetes. It must display daily glucose variability (F19) and clearly illustrate how lifestyle factors such as nutrition, physical activity, and medication intake influence glucose levels over time (S3). As one participant emphasized: “That you basically just hover over the [glucose] curve and then have those [nutrition, activity, medication] data” (Participant 4, focus group 2).

The use of AI to adjust medication based on glucose sensor data was identified as a potential future feature (F20). The focus groups also highlight the importance of wireless transfer of glucose data to the dashboard (I8), synchronization with various glucose monitors (F21), and the option to adjust blood glucose thresholds to allow for individualized targets (F22).

Additional features included displaying how frequently persons with diabetes scan their glucose sensors (F23). This information helps HCPs identify potential issues such as excessive scanning, which could indicate increased stress or anxiety related to glucose monitoring. Notifications with general advice should be sent to the person with diabetes when glucose values are out

XSL•FO
RenderX

of range (F24). Regarding forwarding alerts to HCPs, participants agreed that this should only happen in specific, well-defined cases (F25), with persons with diabetes remaining primarily responsible for acting on their glucose data (I9). Alerts should furthermore be filtered according to predefined clinical protocols to avoid unnecessary notifications (F26).

Requirements for Psychosocial Aspects of Living With T2DM

Four requirements address the role of the social domain in T2DM management, as shown in Table 7. Participants emphasized that both persons with diabetes and HCPs often lack insight into available social support:

> *The patient usually doesn't know where to go.* [Participant 2, focus group 2]
>
> *And neither do we.* [Participant 5, focus group 2]

This shared lack of awareness underscores the need for better integration of social support information in clinical practice. To address this, HCPs suggested incorporating regional social domain information into the dashboard (F27). This would enable providers to more easily access and refer to relevant resources.

Beyond informational access, HCPs also emphasized the direct influence of social circumstances on diabetes management. Stress was specifically highlighted as an essential factor (F28), with the social environment, including work and home life, impacting glucose levels and well-being.

> *Stress factors, social home situation. Just how things are going.* [Participant 2, focus group 1]

Similarly, another participant stressed the relevance of mental well-being as a key aspect of the social domain:

> *And I do ask about mental well-being as well, because it is one of the first aspects to assess within social factors.* [Participant 4, focus group 1]

To integrate these dimensions into the dashboard, HCPs suggested using the Positive Health spider web [45] as a framework (C4). This tool offers a structured way to assess various aspects of well-being, making it easier to identify challenges and provide targeted support. Additionally, HCPs expressed a desire for persons with diabetes to have the ability to add personal notes within the dashboard to contextualize their glucose data, particularly in relation to social or emotional challenges (F29).

Additional Parameter Requirements

HCPs emphasized that T2DM management requires attention to the broader cardiometabolic risk profile. In line with this, HCPs highlighted the need to access up-to-date medication overviews (C5), integration of laboratory results such as hemoglobin A_{1c} levels and blood- and urine cultures (C6), and the visualization of blood pressure and weight (F30), as shown in Table 8: "And furthermore, it would be nice if, across the entire dashboard, you could also include parameters like blood pressure and weight" (Participant 1, focus group 1).

Weight was identified as an important health parameter to monitor, as weight loss is known to reverse the underlying metabolic abnormalities of type 2 diabetes and thereby improve glucose control [46]. To support this process, HCPs noted that weight trends should be displayed over time to enable monitoring of long-term progress (C7). Many people with diabetes also have hypertension. Elevated blood pressure combined with diabetes increases the risk of cardiovascular and renal complications, making routine monitoring highly relevant for clinical decision-making [47].

Sleep was also considered potentially informative, as poor sleep can negatively affect glucose regulation and overall well-being. Tracking indicators such as total sleep time and sleep disturbances (F31) may therefore offer additional insight, although participants differed in how essential they perceived this parameter to be.

Discussion

Principal Findings

This study aimed to identify requirements for a personalized blended dashboard for T2DM care, based on the perspectives of HCPs. The aim was to facilitate informed decision-making with persons with diabetes, enabling more individualized care with personalized goals in a blended care setting. The findings reveal strong support for integrating key lifestyle parameters such as physical activity, dietary intake, glucose values, and psychosocial factors, as well as overall system requirements and additional health-related parameters (eg, medication use and sleep). They underscore the need for an integrated approach to lifestyle monitoring in diabetes care, recognizing the interaction between behaviors such as diet, physical activity, glucose levels, and psychosocial factors. This reflects a growing recognition in the literature that behavioral and emotional contexts play a critical role in diabetes self-management [48-51]. Integrating lifestyle and behavioral data into care underscores the need for shared decision-making tools that facilitate collaboration between patients and HCPs and support personalized diabetes treatment.

Findings in Context

The need for dashboards supporting personalized and blended care is also seen in the literature. Existing shared decision-making tools and/or decision support tools for T2DM address a variety of aspects. Some focus on medication-related decisions, including guidance on drug interactions, dosage adjustments, or therapy changes [52-56]. Others incorporate lifestyle-related data [57-59], including structured forms for diet, physical activity, glucose data, and self-management. However, these tools rarely facilitate the collaborative setting of personalized lifestyle goals with persons with diabetes. For example, the integrated personalized diabetes management system [58] involves persons with diabetes in therapy adjustments based on self-measured blood glucose, but does not integrate diet or physical activity into home monitoring and provides limited support for lifestyle goal setting. Similarly, the Smart Care service offers video consultations, remote monitoring, automated feedback, and educational resources [59] without actively guiding personalized lifestyle decisions. Compared to these existing tools, the requirements identified in our focus groups emphasize offering concrete and personalized lifestyle-related support and the integration of

XSL•FO
RenderX

personal values into decision-making. Two exceptions can be highlighted. MyDiabetesPlan is an online decision aid that supports shared decision-making by assessing the clinical and psychosocial profiles of persons with diabetes, eliciting personal goals, and providing a tailored action plan with concrete strategies to achieve them [60]. However, it does not incorporate home-monitored lifestyle data, limiting its ability to provide truly data-driven lifestyle feedback. Another exception is found in pregnancy care, where a dashboard has been developed for women with T2DM by combining clinical data, lifestyle monitoring, and patient-centered decision support [61]. However, as far as we know, no dashboard currently exists for the broader T2DM population that integrates home-monitored lifestyle behaviors with clinical data. Such a solution would ideally provide patient-centered decision support in a single, personalized tool suitable for use in a blended care pathway, preferably within a single uniform platform that can be used across both primary and secondary care.

International guidelines for the treatment of T2DM emphasize the importance of lifestyle advice, highlighting the crucial role of a healthy lifestyle [19,62,63]. However, translating these recommendations into daily practice remains a challenge, as reflected by our focus groups, where HCPs expressed uncertainty about referrals in the social domain and lacked knowledge in providing personalized nutrition consultations. This highlights the need for clear and accessible dashboard interfaces that integrate clinical guidelines with health data of persons with diabetes. In addition to that, it shows a need for ongoing professional development and training. In the Netherlands, initiatives such as the “Dutch Coalition for Lifestyle in Care” and the “Association for Physicians and Lifestyle” are trying to address this need by promoting structured education and embedding lifestyle medicine more firmly in routine care [64]. At the same time, broader national infrastructure adjustments are required. The integration of lifestyle medicine into existing pathways must be supported by new collaborative structures that enable multiple HCPs to contribute. Such collaboration requires patients to actively consent to sharing their data across providers, and this process is still rarely implemented in current practice [65].

HCPs emphasized that the dashboard must align with their workflow and clinical responsibilities. This reflects a broader challenge in implementing digital tools in routine care, as such tools are most effective when they support rather than disrupt clinical processes [66,67]. Tailoring the information of persons with diabetes to individual needs and integrating features such as alert or notification systems can enhance efficiency and safety of persons with diabetes. At the same time, HCPs stressed that automated alerts should not shift accountability away from persons with diabetes, highlighting the need to balance technological support with responsibility of persons with diabetes.

This study focused on the perspective of HCPs to identify requirements for a dashboard supporting collaborative goal setting and shared decision-making. While this provided important insights into clinical workflows, it must be acknowledged that shared decision-making is inherently bidirectional and requires the active involvement of persons with diabetes. Given that persons with diabetes are expected to monitor their lifestyle and use supportive tools such as continuous glucose monitors or activity trackers, it is essential to incorporate their perspectives to promote self-management and strengthen shared decision-making [68]. Previous work by colleagues from our research group, such as the study by den Braber et al [69], has highlighted the patients’ perspective. This allows us to focus discussions in this study on HCPs, while bringing the different perspectives together again in future work. At the same time, practical considerations such as accessibility, availability, and reimbursement of these tools must be addressed to ensure that sufficient and reliable data can be collected for clinical use [70,71].

Implications for Practice

The findings indicate that strengthening lifestyle care requires integration of lifestyle medicine into routine practice. Moreover, care pathways should be redesigned to center around a lifestyle counter that enables interdisciplinary collaboration and sustainable use of digital tools. Rather than proposing an entirely new pathway, a complementary model could be envisioned, in which a lifestyle counter team supports existing care structures by coordinating persons with diabetes guidance, monitoring, and follow-up. This approach may be particularly valuable in secondary care settings, where persons with diabetes often have completed a lifestyle intervention, such as a Combined Lifestyle Intervention [72]. In such a model, the dashboard would integrate seamlessly into the workflow by guiding persons with diabetes to appropriate professionals, managing alerts, and coordinating follow-up across medical, dietary, psychosocial, and community-based services. To support this approach, the Dutch Coalition for Lifestyle in Care offers information cards that provide practical guidance and resources for implementing Lifestyle Front Office [73,74]. Embedding the dashboard within such a Lifestyle Front Office ensures that lifestyle monitoring becomes a structurally integrated element of diabetes care while reducing workload for individual providers [75].

This study highlights the need for a dashboard that provides insight into personalized blended care for type 2 diabetes, going beyond clinical data alone. As is good practice in any brainstorm situation, and given the fact that the aim of this paper was merely to explore any potential requirements for the dashboard, no requirements were considered “inappropriate” or “unfeasible” in this study. However, we recognize that not all requirements are equally feasible to implement. For example, HCPs suggested a customizable home screen, which is relatively straightforward to implement, whereas more advanced features, such as AI-driven medication adjustment suggestions based on glucose data, may face technical, cost, or ethical constraints. Although medication management did not emerge prominently in the focus groups, integrating medication-related functionalities into future dashboard versions will be important to create a more comprehensive tool for T2DM care. Assessing feasibility and acceptability is therefore an important consideration for future work.

Strengths and Limitations

A key strength of this study is the inclusion of health care providers with diverse professional backgrounds, which

XSL•FO
RenderX

enhanced the comprehensiveness of the findings. Despite this, several limitations should be acknowledged. Given the exploratory nature of this study, where the goal is to general conceptual insights rather than achieve statistical power, using a relatively small number of participants was fitting and justifiable [40]. However, this does imply that the generalizability of the findings is limited. In addition, female participants were overrepresented, which may introduce bias. However, sex-related effects could not be assessed separately from professional role. Moreover, focus group dynamics may have influenced the prioritization of requirements even though researchers prioritized monitoring and guiding the discussion process throughout the focus group to minimize risk. Despite its potential limitations, the study provides valuable insights into the needs and expectations of HCPs regarding a dashboard tailored to the treatment of people with T2DM. By translating these insights into concrete functional and contextual requirements, this study offers a foundation for the development of a dashboard that can strengthen lifestyle support in diabetes care and facilitate more integrated, multidisciplinary collaboration.

Conclusions

This study identified key requirements for a remote patient monitoring dashboard for personalized blended diabetes care, emphasizing the integration of data-driven lifestyle parameters alongside clinical data to support shared decision-making and individualized care. The findings highlight strong support for integrating physical activity, dietary intake, glucose levels, psychosocial factors, and additional parameters (eg, weight and blood pressure) to enable an integrated approach to T2DM management. At the same time, the study underscores both the promise and the challenges of translating lifestyle monitoring into routine diabetes care. This includes the need for clear and accessible visualization of lifestyle data, seamless integration into clinical workflows, and professional training in domains such as nutrition and support and psychosocial referral options. A meaningful dashboard must therefore be clinically relevant and user-friendly but also embedded within broader structural changes in health care practice and professional education. This includes multidisciplinary access and alignment with professional guidelines. Addressing both functional and contextual requirements, alongside the development of fitting care pathways, is essential to ensure that a dashboard truly supports HCPs in delivering integrated and personalized lifestyle care.

Acknowledgments

The authors would like to thank all health care professionals who participated in the focus groups and shared their insights and experiences. The authors used OpenAI's GPT-5 (via ChatGPT) for language verification and overall readability enhancement.

Funding

This project was funded by the Dutch Diabetes Research Foundation (project 2022.30.001) and by a grant from the municipality of Almelo. Both entities were not involved in the design, collection, analysis or interpretation of the data, in the writing of the manuscript, or in the decision to publish this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the sensitive nature of the focus group transcripts and to protect participant confidentiality. Anonymized excerpts relevant to the findings are available upon reasonable request from the corresponding author.

Authors' Contributions

CL, NB-J, and AM contributed to the development and design of the study. Data collection was carried out by CL and EH, with CL also responsible for data extraction in collaboration with NB-J. CL primarily drafted the manuscript and collaborated closely with NB-J and AM. GDL, EH, AM, and NB-J provided ongoing feedback on the Methods, Results, and Discussion sections. All authors critically revised the manuscript, contributed to its finalization, and approved the definitive version for publication.

Conflicts of Interest

None declared.

References

1. Kharroubi AT, Darwish HM. Diabetes mellitus: the epidemic of the century. World J Diabetes 2015 Jun 25;6(6):850-867. [doi: 10.4239/wjd.v6.i6.850] [Medline: 26131326]
2. Diabetes in cijfers. Diabetes Fonds. URL: https://www.diabetesfonds.nl/over-diabetes/wat-is-diabetes/diabetes-in-cijfers [accessed 2025-09-16]
3. Diabetes mellitus | leeftijd en geslacht. Volksgezondheid en Zorg. URL: https://www.vzinfo.nl/diabetes-mellitus/leeftijd-en-geslacht [accessed 2025-09-16]

XSL•FO
RenderX

4. Zimmet PZ, Magliano DJ, Herman WH, Shaw JE. Diabetes: a 21st century challenge. Lancet Diabetes Endocrinol 2014 Jan;2(1):56-64. [doi: 10.1016/S2213-8587(13)70112-8] [Medline: 24622669]
5. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018 Feb;14(2):88-98. [doi: 10.1038/nrendo.2017.151] [Medline: 29219149]
6. Lean MEJ, Leslie WS, Barnes AC, et al. Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. Lancet Diabetes Endocrinol 2019 May;7(5):344-355. [doi: 10.1016/S2213-8587(19)30068-3]
7. Ried-Larsen M, Johansen MY, MacDonald CS, et al. Type 2 diabetes remission 1 year after an intensive lifestyle intervention: a secondary analysis of a randomized clinical trial. Diabetes Obes Metab 2019 Oct;21(10):2257-2266. [doi: 10.1111/dom.13802] [Medline: 31168922]
8. Marín-Peñalver JJ, Martín-Timón I, Sevillano-Collantes C, Del Cañizo-Gómez FJ. Update on the treatment of type 2 diabetes mellitus. World J Diabetes 2016 Sep 15;7(17):354-395. [doi: 10.4239/wjd.v7.i17.354] [Medline: 27660695]
9. Lancel M, van Veen MM, Verbeek I. Sleep: the foundation of a healthy lifestyle [Slaap: de basis voor een gezonde leefstijl]. Tijdschr Psychiatr 2020 Nov 12;62(11):949-954. [Medline: 33443745]
10. Wentink-van de Laar J, Weerkamp-de Boer J, Aantjes C. Zorgstandaard leefstijl. Nurse Academy GGZ. 2024. URL: https://akwaggz.nl/wp-content/uploads/2024/06/Zorgstandaard-Leefstijl-Nurse-Academy-GGZ-nummer-1-2024.pdf [accessed 2026-03-22]
11. Beweegrichtlijnen 2017. Gezondheidsraad. URL: https://www.gezondheidsraad.nl/documenten/2017/08/22/beweegrichtlijnen-2017 [accessed 2025-09-19]
12. Sport en bewegen in cijfers. Beweegrichtlijnen. URL: https://www.sportenbewegenincijfers.nl/kernindicatoren/beweegrichtlijnen [accessed 2026-03-22]
13. Oosterom N, Gant CM, Ruiterkamp N, et al. Physical activity in patients with type 2 diabetes: the case for objective measurement in routine clinical care. Diabetes Care 2018 Apr;41(4):e50-e51. [doi: 10.2337/dc17-2041] [Medline: 29432126]
14. Gant CM, Binnenmars SH, Berg EVD, Bakker SJL, Navis G, Laverman GD. Integrated assessment of pharmacological and nutritional cardiovascular risk management: blood pressure control in the DIAbetes and LifEstyle Cohort Twente (DIALECT). Nutrients 2017 Jul 6;9(7):709. [doi: 10.3390/nu9070709] [Medline: 28684676]
15. Gant CM, Binnenmars SH, Harmelink M, et al. Real-life achievement of lipid-lowering treatment targets in the DIAbetes and LifEstyle Cohort Twente: systemic assessment of pharmacological and nutritional factors. Nutr Diabetes 2018;8(1). [doi: 10.1038/s41387-018-0028-y]
16. Jalving AC, Gant CM, Binnenmars SH, et al. Glycaemic control in the diabetes and Lifestyle Cohort Twente: a cross-sectional assessment of lifestyle and pharmacological management on Hba1c target achievement. Diabetes Obes Metab 2018 Oct;20(10):2494-2499. [doi: 10.1111/dom.13399] [Medline: 29862616]
17. Leefstijl in richtlijnen. Coalitie Leefstijl in de Zorg. 2024. URL: https://leefstijlcoalitie.nl/wp-content/uploads/2024/05/Leefstijl-in-Richtlijnen.pdf [accessed 2026-03-16]
18. American Diabetes Association Professional Practice Committee. Introduction and methodology: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S1-S5. [doi: 10.2337/dc25-SINT] [Medline: 39651982]
19. Rosenfeld RM, Grega ML, Gulati M. Lifestyle interventions for treatment and remission of type 2 diabetes and prediabetes in adults: implications for clinicians. Am J Lifestyle Med 2025 Mar 25:15598276251325802. [doi: 10.1177/15598276251325802] [Medline: 40161282]
20. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011 Jul;93(1):1-9. [doi: 10.1016/j.diabres.2011.02.002] [Medline: 21382643]
21. Wolpert HA, Anderson BJ. Management of diabetes: are doctors framing the benefits from the wrong perspective? BMJ 2001 Oct 27;323(7319):994-996. [doi: 10.1136/bmj.323.7319.994] [Medline: 11679393]
22. Pot CHM, Hietbrink EAG, Laverman GD. Implementing lifestyle counselling into secondary type 2 diabetes care: perspectives from diabetes nurses. Nurs Open 2025 Apr;12(4):e70183. [doi: 10.1002/nop2.70183] [Medline: 40223769]
23. Singh B, Ahmed M, Staiano AE, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 5;7(1):179. [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
24. Davis A, Sweigart R, Ellis R. A systematic review of tailored mHealth interventions for physical activity promotion among adults. Transl Behav Med 2020 Oct 12;10(5):1221-1232. [doi: 10.1093/tbm/ibz190] [Medline: 33044542]
25. Raber M, Liao Y, Rara A, et al. A systematic review of the use of dietary self-monitoring in behavioural weight loss interventions: delivery, intensity and effectiveness. Public Health Nutr 2021 Dec;24(17):5885-5913. [doi: 10.1017/S136898002100358X] [Medline: 34412727]
26. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
27. Kong SY, Cho MK. Effects of continuous glucose monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis. Healthcare (Basel) 2024 Feb 29;12(5):571. [doi: 10.3390/healthcare12050571] [Medline: 38470682]

XSL•FO
RenderX

28. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. NPJ Digit Med 2024 Jul 18;7(1):192. [doi: 10.1038/s41746-024-01182-w] [Medline: 39025937]
29. Dagliati A, Sacchi L, Tibollo V, et al. A dashboard-based system for supporting diabetes care. J Am Med Inform Assoc 2018 May 1;25(5):538-547. [doi: 10.1093/jamia/ocx159]
30. van Beurden SB, Simmons SI, Tang JCH, Mewse AJ, Abraham C, Greaves CJ. Informing the development of online weight management interventions: a qualitative investigation of primary care patient perceptions. BMC Obes 2018;5(1):7. [doi: 10.1186/s40608-018-0184-6] [Medline: 29449953]
31. Lyles CR, Harris LT, Le T, et al. Qualitative evaluation of a mobile phone and web-based collaborative care intervention for patients with type 2 diabetes. Diabetes Technol Ther 2011 May;13(5):563-569. [doi: 10.1089/dia.2010.0200] [Medline: 21406018]
32. Huygens MWJ, Vermeulen J, Swinkels ICS, Friele RD, van Schayck OCP, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 8;16(1):232. [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
33. Hohberg V, Fuchs R, Gerber M, Künzler D, Paganini S, Faude O. Blended care interventions to promote physical activity: a systematic review of randomized controlled trials. Sports Med Open 2022 Jul 30;8(1):100. [doi: 10.1186/s40798-022-00489-w] [Medline: 35907158]
34. Lau Y, Chee DGH, Chow XP, Cheng LJ, Wong SN. Personalised eHealth interventions in adults with overweight and obesity: a systematic review and meta-analysis of randomised controlled trials. Prev Med 2020 Mar;132:106001. [doi: 10.1016/j.ypmed.2020.106001] [Medline: 31991155]
35. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024 Jul 22;26:e55757. [doi: 10.2196/55757] [Medline: 39037772]
36. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
37. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e35479. [doi: 10.2196/35479] [Medline: 35394443]
38. Abbasgholizadeh Rahimi S, Menear M, Robitaille H, Légaré F. Are mobile health applications useful for supporting shared decision making in diagnostic and treatment decisions? Glob Health Action 2017 Jun;10(Suppl 3):1332259. [doi: 10.1080/16549716.2017.1332259] [Medline: 28838306]
39. Raats I. Een praktische handleiding voor de organisatie, begeleiding en analyse van focusgroepen. Handleiding Focusgroepen. URL: https://participatiekompas.nl/media/pdf/handleiding_focusgroepen_2023-toegankelijk.pdf [accessed 2026-03-22]
40. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field methods 2017 Feb;29(1):3-22. [doi: 10.1177/1525822X16639015]
41. Kravchenko T, Bogdanova T, Shevgunov T. In: Silhavy R, editor. Ranking Requirements Using MoSCoW Methodology in Practice: Springer International Publishing; 2022:188-199.
42. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21. [doi: 10.2196/resprot.2547] [Medline: 23796508]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Widya I, Bults R, Wijk R, et al. In: Berry D, Franch X, editors. Requirements for a Nutrition Education Demonstrator: Springer; 2011:48-53.
45. Home. MijnPositieveGezondheid.nl. URL: https://mijnpositievegezondheid.nl/ [accessed 2026-01-16]
46. Lingvay I, Sumithran P, Cohen RV, le Roux CW. Obesity management as a primary treatment goal for type 2 diabetes: time to reframe the conversation. The Lancet 2022 Jan;399(10322):394-405. [doi: 10.1016/S0140-6736(21)01919-X]
47. Hoge bloeddruk. Diabetes Fonds. URL: https://www.diabetesfonds.nl/over-diabetes/dagelijks-leven/hoge-bloeddruk-hypertensie [accessed 2026-01-16]
48. Davies M. Psychological aspects of diabetes management. Medicine (Abingdon) 2022 Nov;50(11):749-751. [doi: 10.1016/j.mpmed.2022.08.011]
49. Halliday JA, Speight J, Bennet A, Beeney LJ, Hendrieckx C. The diabetes and emotional health handbook and toolkit for health professionals supporting adults with type 1 and type 2 diabetes: formative evaluation. JMIR Form Res 2020 Feb 21;4(2):e15007. [doi: 10.2196/15007] [Medline: 32130112]
50. Kalra S, Jena BN, Yeravdekar R. Emotional and psychological needs of people with diabetes. Indian J Endocr Metab 2018;22(5):696. [doi: 10.4103/ijem.IJEM_579_17]
51. Standpunt ‘Remissie als een behandeluitkomst bij volwassenen met DM2. .
52. Kunneman M, Branda ME, Ridgeway JL, et al. Making sense of diabetes medication decisions: a mixed methods cluster randomized trial using a conversation aid intervention. Endocrine 2022 Feb;75(2):377-391. [doi: 10.1007/s12020-021-02861-4]

XSL•FO
RenderX

53. Karagiannis T, Liakos A, Branda ME, et al. Use of the Diabetes Medication Choice Decision Aid in patients with type 2 diabetes in Greece: a cluster randomised trial. BMJ Open 2016 Nov 14;6(11):e012185. [doi: 10.1136/bmjopen-2016-012185] [Medline: 28186933]
54. Bailey RA, Shillington AC, Harshaw Q, Funnell MM, VanWingen J, Col NF. Changing patients' treatment preferences and values with a decision aid for type 2 diabetes mellitus: results from the treatment arm of a randomized controlled trial. Diabetes Ther 2018 Apr;9(2):803-814. [doi: 10.1007/s13300-018-0391-7] [Medline: 29536425]
55. Hsu WC, Lau KHK, Huang R, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016 Feb;18(2):59-67. [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
56. Rickard JP, Zhou S, Kumar N, Jackson HB, Chen K. Creating a dashboard for oversight of a comprehensive medication management program for patients with type 2 diabetes in a large municipal health care system in the United States. J Am Coll Clin Pharm 2025 May;8(5):358-366. [doi: 10.1002/jac5.70026]
57. Welch G, Zagarins SE, Santiago-Kelly P, et al. An internet-based diabetes management platform improves team care and outcomes in an urban Latino population. Diabetes Care 2015 Apr;38(4):561-567. [doi: 10.2337/dc14-1412] [Medline: 25633661]
58. Heinemann L, Schramm W, Koenig H, et al. Benefit of digital tools used for integrated personalized diabetes management: results from the PDM-ProValue study program. J Diabetes Sci Technol 2020 Mar;14(2):240-249. [doi: 10.1177/1932296819867686] [Medline: 31378074]
59. Jeong JY, Jeon JH, Bae KH, et al. Smart care based on telemonitoring and telemedicine for type 2 diabetes care: multi-center randomized controlled trial. Telemed J E Health 2018 Aug;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]
60. Yu CH, Medleg F, Choi D, et al. Integrating shared decision-making into primary care: lessons learned from a multi-centre feasibility randomized controlled trial. BMC Med Inform Decis Mak 2021 Nov 22;21(1):323. [doi: 10.1186/s12911-021-01673-w] [Medline: 34809626]
61. Fareed N, Swoboda C, Singh P, et al. Developing and testing an integrated patient mHealth and provider dashboard application system for type 2 diabetes management among Medicaid-enrolled pregnant individuals based on a user-centered approach: mixed-methods study. Digit Health 2023;9:20552076221144181. [doi: 10.1177/20552076221144181] [Medline: 36644662]
62. American Diabetes Association. 5. Lifestyle Management: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S46-S60. [doi: 10.2337/dc19-S005] [Medline: 30559231]
63. Aas AM, Axelsen M, Churuangsuk C. Evidence-based European recommendations for the dietary management of diabetes. Diabetologia 2023 Jun;66(6):965-985. [doi: 10.1007/s00125-023-05894-8] [Medline: 37069434]
64. Leefstijl Coalitie in de Zorg. URL: https://leefstijlcoalitie.nl/
65. Keuper J, Hek K, van Tuyl LHD, Batenburg R, Verheij R. Determinants of consent for electronic health information exchange: an observational retrospective study. Health Res Policy Syst 2025 Aug 7;23(1):103. [doi: 10.1186/s12961-025-01377-x] [Medline: 40775343]
66. Lansink C, Middelweerd A, Laverman GD, Jong ND. Barriers and facilitators for implementation of lifestyle related blended care interventions for overweight, obesity and type 2 diabetes mellitus: a scoping review. Implement Sci Commun 2026 Mar 19. [doi: 10.1186/s43058-026-00904-3] [Medline: 41857778]
67. Peyroteo M, Ferreira IA, Elvas LB, Ferreira JC, Lapão LV. Remote monitoring systems for patients with chronic diseases in primary health care: systematic review. JMIR Mhealth Uhealth 2021 Dec 21;9(12):e28285. [doi: 10.2196/28285] [Medline: 34932000]
68. Morken V, Perry LM, Coughlin A, et al. Implementation of a co-design strategy to develop a dashboard to support shared decision making in advanced cancer and chronic kidney disease. J Clin Med 2024 Jul 17;13(14):4178. [doi: 10.3390/jcm13144178] [Medline: 39064218]
69. den Braber N, Vollenbroek-Hutten MMR, Oosterwijk MM, et al. Requirements of an application to monitor diet, physical activity and glucose values in patients with type 2 diabetes: the diameter. Nutrients 2019 Feb 15;11(2):409. [doi: 10.3390/nu11020409] [Medline: 30781348]
70. Hulpmiddelen bij diabetes (Zvw). Zorginstituut Nederland. URL: https://www.zorginstituutnederland.nl/verzekerde-zorg/h/hulpmiddelen-bij-diabetes [accessed 2026-01-16]
71. Vergoeding van glucosesensoren. Diabetesvereniging Nederland. URL: https://www.dvn.nl/jouw-belangen/vergoedingen/vergoeding-glucosesensoren [accessed 2026-01-16]
72. Gecombineerde leefstijlinterventie (GLI) (Zvw). Zorginstituut Nederland. URL: https://www.zorginstituutnederland.nl/verzekerde-zorg/g/gecombineerde-leefstijlinterventie-gli-zvw [accessed 2026-01-16]
73. Hoe organiseer ik een leefstijl(zorg)loket voor patiënten in of vanuit het ziekenhuis? Coalitie Leefstijl in De Zorg. URL: https://www.nvz-ziekenhuizen.nl/sites/default/files/2024-05/Hoe-organiseer-ik-een-leefstijlzorgloket-voor-patienten-in-of-vanuit-het-ziekenhuis-1.pdf [accessed 2026-03-16]
74. Implementatiekaart netwerk: van spreekkamer naar leefstijlinterventie. Nederlandse Verening van Ziekenhuizen Federatie Medisch Specialisten. URL: https://demedischspecialist.nl/sites/default/files/2022-03/implementatiekaart_van_spreekkamer_naar_leefstijlinterventie.pdf [accessed 2026-03-16]

XSL•FO
RenderX

75. Leefstijl(zorg)loketten. Coalitie Leefstijl in de Zorg. URL: https://leefstijlcoalitie.nl/implementatie/leefstijlzorgloketten/ [accessed 2026-01-16]

Abbreviations

AI: artificial intelligence
HCP: health care professional
MoSCoW: must/should/could/will not have
T2DM: type 2 diabetes mellitus

Edited by J Lu, S Li; submitted 30.Sep.2025; peer-reviewed by J Ni, Y Wang, Y Mo; revised version received 30.Jan.2026; accepted 08.Feb.2026; published 30.Mar.2026.

Please cite as:
Lansink C, Beerlage-de Jong N, Hietbrink E, Middelweerd A, Laverman GD
Exploring the Needs of Health Professionals for a Type 2 Diabetes Remote Patient Monitoring Dashboard for Personalized Care: Focus Group Study
JMIR Diabetes 2026;11:e84894
URL: https://diabetes.jmir.org/2026/1/e84894
doi:10.2196/84894

© Chiara Lansink, Nienke Beerlage-de Jong, Eclaire Hietbrink, Anouk Middelweerd, Gozewijn Dirk Laverman. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bridging the Digital Divide in Diabetes Care With an AI-Powered Health Coaching Intervention: Qualitative Formative Study

Tejossy John[1,2], BPT, MPH; Sangeetha Mohanraj[3], MSc, PhD; Tanjila Nawshin[4], MS, MBBS; Tapan Mehta[2], MSEE, PhD; Mohanraj Thirumalai[5], MS, MEng, PhD

[1]Department of Orthopedic Surgery, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Department of Family and Community Medicine, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Research Collaborative, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[5]Division of General Internal Medicine and Population Health, Heersink School of Medicine, University of Alabama at Birmingham, 701 19th Street South, ALGEN 580, Birmingham, AL, United States

Corresponding Author:
Mohanraj Thirumalai, MS, MEng, PhD
Division of General Internal Medicine and Population Health, Heersink School of Medicine, University of Alabama at Birmingham, 701 19th Street South, ALGEN 580, Birmingham, AL, United States

Abstract

Background: Despite advancements in digital health coaching (DHC) for type 2 diabetes self-management, a persistent digital divide limits access among underserved populations, necessitating low-tech, telephone-based alternatives.

Objective: This formative qualitative study explored the lived experiences of type 2 diabetes mellitus (T2DM) self-management challenges among adults in Alabama and elicited preferences for telephone-based DHC to inform intervention optimization.

Methods: This study was part of a larger National Institutes of Health clinical trial (5R01DK129378) that examined the feasibility and needs of a telephone-based DHC intervention for self-management of T2DM. The study employed a qualitative, phenomenological approach to assess patients' needs surrounding digital coaching. Twelve patients diagnosed with T2DM were recruited between August and December 2022. Data collection involved in-depth, semistructured interviews conducted via telephone calls or secure Zoom sessions. All interviews were audio-recorded, professionally transcribed, and verified for accuracy. Two independent coders analyzed the data using NVivo software version 14 Plus.

Results: Two themes emerged: (1) multifaceted barriers to diabetes self-management spanning behavioral (dietary adherence and nutritional gaps), physical/environmental (mobility limitations and weather constraints), and structural domains (medication shortages, insurance coverage limiting continuous glucose monitoring); (2) expectations for telephone-based DHC, including targeted education (meal planning, exercise adaptations, and medication effects), health coaches as accountability partners, daily reminders and check-ins via texts/calls, and incentives as extrinsic motivators to adherence. Insights from the study directly shaped the intervention components of a pilot feasibility trial (NCT05344859).

Conclusions: Key challenges in day-to-day T2DM management included diet, physical activity, medication unavailability, and lack of insurance coverage for continuous glucose monitors. Expectations for a potential telephone-based DHC intervention included targeted education on diet, exercise, and medication effects; perceptions of health coaches as accountability partners; reminders via texts/calls as beneficial; and monetary incentives as extrinsic motivators for adherence to the intervention.

(JMIR Diabetes 2026;11:e81441) doi:10.2196/81441

KEYWORDS

type 2 diabetes; digital health coaching; patient education; diabetes self-management; patient experience; qualitative research; phenomenology; diabetes technology; expectations

Introduction

Type 2 diabetes mellitus (T2DM), a chronic metabolic disorder characterized by insulin resistance and impaired glucose regulation, presents a significant public health concern, imposing a substantial burden on individual well-being and the health care system in the United States. According to the National Diabetes Statistics Report published by the Centers for Disease Control and Prevention, approximately 38.4 million individuals are currently struggling to manage uncontrolled diabetes. Concurrently, an estimated 8.7 million adults remain undiagnosed with the condition [1].

XSL•FO
RenderX

Primary care physicians (PCPs) play a central role in diabetes management, delivering care to approximately 90% of individuals diagnosed with the condition amid a nationwide shortage of board-certified endocrinologists [2-4]. PCPs assume multifaceted responsibilities for early detection, routine monitoring, treatment planning, and lifestyle counseling [5,6]. However, high patient volumes and limited consultation time contribute to a substantial workload burden. The median primary care visit lasts only 15.7 minutes and typically addresses 6 distinct clinical issues [7], constraining the capacity to deliver comprehensive diabetes care [8-10].

People with T2DM face numerous daily challenges, including a significant need for lifestyle modification [11,12], consistent medication adherence [11], and psychosocial difficulties [13,14], adapting to physical symptoms and management of comorbidities [15-17]. Digital health coaching (DHC) emerged as an evidence-based intervention to alleviate the burden on PCPs and enhance patient outcomes. DHC applies behavioral science, clinical strategies, and skillful conversation to promote sustainable health behavior change, thereby improving diabetes self-management, fostering lifestyle changes, and enhancing clinical outcomes [18-20].

Over the past decade, as smartphone use has surged dramatically [21], this proliferation has paralleled a marked rise in smartphone-based and technology-enabled remote health coaching interventions, shifting traditional care delivery toward scalable, patient-centered models that leverage mobile apps and telehealth platforms. Numerous studies have substantiated the efficacy of digitally powered health coaching in diabetes management, consistently demonstrating clinically meaningful reductions in HbA_{1c} levels [22-26]. Such evidence underscores the potential of digital coaching to address gaps in primary care capacity while empowering patients with real-time support and personalized guidance; however, a digital divide limits equitable access for underserved populations.

However, the penetration of digital technologies, particularly smartphones, continues to expand across the United States [27]. This growth remains uneven and is accompanied by persistent disparities stratified by age, gender, education, and income levels [27]. Beyond material access, such as internet connectivity and smartphone access, many individuals face a second level of digital divide in the skills needed to use these technologies effectively. Trends reported by Hong and Cho [28] on older adults and digital health interventions highlight these challenges and recommend developing senior-friendly resources to overcome age-related barriers to technology adoption and usability. Similarly, a study by Kumar et al [29] highlighted that health interventions that use advanced mobile features may widen disparities by excluding those with low mobile literacy. Simple mobile phone capabilities, such as text messaging, require significant health literacy and have previously affected enrollment [30] and adoption [31] of texting-based interventions.

To bridge the digital divide, we proposed a study design—an intervention delivered via the simplest widely available technology, a telephone call—without requiring internet access, smartphones, or digital skills, thereby ensuring broad inclusivity for underserved populations while remaining fully compatible with modern smartphones for those who prefer or possess them. This exploratory qualitative study was structured as a comprehensive needs assessment to understand patients' priorities, unmet needs, and expectations regarding a telephone-based diabetes self-management intervention. By systematically gathering qualitative insights from a small sample of the target population, we sought to ensure that the intervention design aligned with real-world patient experiences and preferences. This patient-centered approach directly guided the design and development of the interactive voice response (IVR)–supported telephone coaching intervention and was conducted as a formative substudy within the larger NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases)-funded clinical trial (5R01DK129378), which evaluated the feasibility and effectiveness of an AI-powered telephone-based DHC intervention for T2DM management among adults in the Deep South [32].

This study aims to expand the foundational knowledge about patients' lived experiences with T2DM by understanding two key areas: (1) self-management burden, including daily challenges encountered during T2DM self-management, and (2) expectations from telephone-based interventions, including participants' perceptions and preferences for a future DHC platform delivered via phone and compatible technologies. Insights from this formative work were used to develop a user-centric, culturally tailored telephone and IVR-based health-coaching intervention, subsequently operationalized in the GODART (Gamified Optimized Diabetes Management with Artificial Intelligence–Powered Rural Telehealth) IVR tracking and goal-setting system, optimized for adults with T2DM, and designed to be accessible to those facing technological or socioeconomic barriers in the Deep South of the United States.

Methods

Study Design

This exploratory qualitative study employed semistructured interviews and phenomenological thematic analysis as a formative needs assessment to understand patients' challenges and expectations for a telephone-based DHC intervention for T2DM self-management. Semistructured interviews were conducted with a purposive sample of patients with T2DM. The resulting data were analyzed using rigorous, systematic qualitative methods, including thematic analysis and intercoder agreement assessment [33].

Recruitment and Sample

Purposive sampling was conducted among the residents of the state of Alabama who were (1) clinically diagnosed with T2DM, (2) aged 18 years or older, (3) able to converse in English, and (4) had access to a phone. Patient recruitment occurred through the University of Alabama at Birmingham (UAB) Health System. Clinical staff approached patients who visited the Family and Community Medicine Clinic to determine their willingness to participate in a semistructured qualitative interview. Recruitment for this study took place between August and December 2022.

XSL•FO
RenderX

Patients who expressed interest were contacted over the phone by the research team to confirm eligibility and obtain verbal consent. Interviews were then scheduled over the telephone or in a closed Zoom conferencing room for approximately 45 minutes. During the interview, the study staff followed the semistructured interview guide to elicit and record participant responses (Multimedia Appendix 1). The purpose of the interview was to gain an in-depth understanding of the lived experiences and challenges of patients with T2DM living in Alabama, as well as their expectations regarding a telephone-based DHC intervention. Afterward, the audio recordings were transcribed via a professional transcription service, Rev, Inc [34], and thematic analysis was conducted on the resulting transcripts. Data saturation was achieved after 12 semistructured interviews, as no new themes or insights emerged from the last 2 participants. This sample size aligns with established qualitative phenomenological research standards and captures a comprehensive range of patient experiences and expectations related to T2DM management and DHC in this population [35].

Ethical Considerations

The study was approved as exempt by the UAB Institutional Review Board under protocol number IRB-300008277. Verbal informed consent was obtained from all participants during the screening call. To protect privacy and ensure confidentiality, all participant data were deidentified. Participants were compensated US $25 for participating in the interview.

Data Analysis

Analytic Approach

Our basic data analytic approach was exploratory and content-driven [36]. The rationale for choosing an exploratory approach was that it provided an opportunity to inductively expand the knowledge base on patients' experiences with T2DM, the challenges they face in managing the disease, and patients' expectations of a telephone-based DHC program for managing T2DM [37]. The specific approach to analysis was phenomenology because this approach uses people's subjective experiences to gain an in-depth understanding of a phenomenon [38]. Since we explored the challenges patients face in managing T2DM and how those lived experiences shaped their expectations surrounding a telephone-based DHC program, the phenomenological approach appeared to be a good fit for the study. We followed the data analysis steps outlined in the Creswell and Creswell [33] guidelines. We first organized and prepared the transcripts for analysis by transcribing the audio-recorded files.

Development of Codebook

After an initial immersion in the transcripts, the 2 coders independently read all transcripts multiple times and wrote analytic memos to capture their initial impressions. We conducted both the first and second cycles of coding.

First-cycle coding was conducted line-by-line using a combination of descriptive codes (summarizing segments of data) and in vivo codes (participants' own words). Each transcript was initially coded independently by 2 researchers to identify participants' concrete experiences with diabetes self-management and expectations for telephone-based coaching.

After coding the first set of transcripts, the coders met to compare their code lists, collapsed overlapping codes, and agreed on common code labels and definitions, which formed the first version of the codebook. This codebook was iteratively refined as additional transcripts were coded; new codes were added when data segments did not fit existing codes, while similar codes were merged or refined to improve clarity and consistency. Each codebook revision was discussed and agreed upon by both coders, and discrepancies were resolved through discussion, with a third qualitative expert consulted when consensus could not be reached.

In the second coding cycle, the team used pattern coding to group the first cycle of codes into higher-order categories or themes that reflected broader concepts (eg, combining multiple diet-related codes into a broader category reflecting diet management) [39].

These categories were then systematically compared across transcripts to identify recurring patterns and relationships. These categories were further abstracted into themes and subthemes that described patterns across participants. Candidate themes were continuously checked against the coded data and raw transcripts to ensure they were grounded in participants' accounts. Negative or divergent cases that did not fit initial interpretations were examined. Themes were finalized only when all coders agreed that they comprehensively captured the dataset's range of perspectives.

Intercoder Reliability and Rigor

Two researchers independently coded all transcripts, and assistance from a third coder was sought when the first 2 coders could not reach an agreement on coding. The intercoder agreement level was calculated using the Huberman method [39], and an agreement level of >90% was observed. NVivo software version 12 Plus (Lumivero) was used to organize, code, and categorize the information obtained from the transcripts.

A substantial amount of time was spent on the analysis, and expert help was sought when necessary. Deidentified data were used throughout the analysis. An analytic perspective was maintained throughout the analysis, using deidentified data while remaining empathetic to participants' life experiences [40]. Opinions from all participants were given equal weight when interpreting the data to avoid selective interpretation [40]. It was ensured that the data remained fully confidential, and only the researchers involved in this study had access to it [39].

Research Team and Positionality

The analytical team consisted of 2 UAB-based researchers, one with expertise in public health and the other in health services research with prior qualitative research experience, who collaboratively coded the transcripts. Both researchers had long-standing experience working with adults with T2DM in the Deep South, including predominantly African American, socioeconomically disadvantaged populations receiving care within the UAB Health System [32,41,42]. To ensure credible analysis, a third qualitative expert, a health service researcher

XSL•FO
RenderX

with extensive experience in qualitative studies, was consulted in case of coding disagreement or interpretive uncertainty.

Results

Participants and Eligibility Criteria

A total of 12 participants completed semistructured interviews for this study. Their demographic characteristics are summarized in Table 1. Participants were predominantly female (n=10, 83.33%) and African American (n=10, 83.33%), with a mean age of 54.6 (SD 12.39) years, and most had lived with diabetes for 5 to 10 years (n=7, 58.56%). Most participants (n=7, 58.33%) had a high school diploma or equivalency, 3 (25%) had less than high school education, and 2 (16.66%) held bachelor's degrees. The patient Area Deprivation Index (ADI) was categorized as low (ADI value 1 - 3), medium (ADI value 4 - 7), and high (ADI value 7 - 10). The distribution across ADI categories showed that 6 (50%) participants resided in high-deprivation areas, 5 (41.67%) in medium-deprivation areas, and 1 (8.33%) in low-deprivation areas.

Table . Demographic information (N=12).

Patient	Age (y)	Gender	Race	Education	Duration since diagnosis (y)	Area Deprivation Index
P1	59	Female	African American	Less than high school degree	18	7
P2	67	Female	White	High school graduate or GED[a]	1.3	6
P3	73	Female	African American	High school graduate or GED	11	7
P4	32	Male	African American	High school graduate or GED	5	9
P5	53	Female	African American	Bachelor's degree	3	1
P6	67	Male	African American	Less than high school degree	5	10
P7	45	Female	African American	High school graduate or GED	4	10
P8	64	Female	African American	High school graduate or GED	7	5
P9	34	Female	African American	High school graduate or GED	9	5
P10	43	Female	African American	High school graduate or GED	10	8
P11	57	Female	African American	Bachelor's degree	10	9
P12	38	Female	White and Hispanic	Less than high school degree	8	9

[a]GED: General Educational Development.

All participants had confirmed a diagnosis of T2DM for at least 1 year, ranging between 15 months and 18 years. Each participant self-reported multiple comorbid conditions, including, but not limited to, circulatory issues (eg, hypertension), metabolic system disorders (eg, hypercholesterolemia), and diagnoses related to their mental health (eg, anxiety disorder).

Some other reported comorbidities included respiratory conditions (eg, asthma and chronic obstructive pulmonary disease), liver conditions (not specified), allergies, anemia, migraines, sleep difficulties, etc.

Two themes, each with several subthemes, emerged from the analysis of 12 transcripts. The first theme demonstrated the multifaceted challenges participants face in diabetes self-management, and the second theme explored the participants' anticipation of a telephone-based DHC intervention for diabetes management. Each theme and its related subthemes are described in detail in the following section.

Theme 1. Challenges to Diabetes Self-Management

Theme Overview

Diabetes self-management presents persistent, multifaceted challenges that test patients' knowledge, resources, and resilience daily. Participants in this study expressed challenges across four key domains, including (1) managing diet, (2) physical activity, (3) medication access barriers, and (4) blood glucose monitoring (Table 2).

XSL•FO
RenderX

Table . Participant-reported challenges to diabetes self-management.

Domain	Subthemes	Description	Quotes
Behavioral	Managing diet	Struggles with dietary adherence, cravings, and lack of nutritional literacy	• "Well, when I first was told that I had diabetes, it was kind of hard for me to eat different foods, because I was scared, I was eating the wrong thing… I don't know how to go in the grocery store and read on the label, cholesterol this, or this much this, that, and another…" (P8) • "Basically, eating sweet stuff.. that's very hard, but I try to manage it. If I do eat it, I just take a little bit of more insulin." (P9) • "A lot of times I don't feel like cooking anything, so I'll just maybe eat some cereal or maybe drink a Boost in the morning." (P3) • "Yes. It's very difficult to me. When you work, that's a little harder because you can't always eat six meals a day." (P5) • "My challenges that I have, is years of eating a certain way and doing a certain thing, and your mindset of thinking that you're Superman and nothing can hurt you. Now I'm finding out that I have high blood pressure and diabetes, so I'm having to make a life change.." (P4)
Physical and environmental	Mobility constraints	Functional and environmental limitations that limit physical activity	• "Well, problems I have been having since I had diabetes.. My leg's been hurting…and it messed with my walking. It messed up my legs up. I can't hardly walk sometimes." (P7) • "I have challenge with my ankles and stuff swelling up, and I sweat a lot.. So, it is difficult to walk" (P1) • I do. I try to walk. I haven't been doing that lately. (Because it is cold outside) (P11)

XSL•FO
RenderX

Domain	Subthemes	Description	Quotes
Structural	Medication and monitoring access	Systemic barriers, including unaffordable copays, medication stock-outs, and lack of insurance for CGM[a] devices	• "It goes up and down (referring to blood sugar levels). Because I don't take no medicines... Since I was taking the shot once a week, and they went up on my diabetes (medication costs)….they went up on it, and my insurance did… and I couldn't afford what they went to, cause I was paying a US $40 copay, and it went up to 196." (P6) • "It's not that I'm forgetting. For the past three months, I've been having trouble getting my medication. Trulicity is never in stock, so she (primary care provider) changed it to, I don't want to get the name wrong, but it starts with an O (referring to a new medication)… and that's not in stock either. So the last doctor visit I had, she changed it to yet another medication and it is not covered on Medicaid… so I reached out to her on my portal to let her know… I don't know what I'm going to receive now." (P10) • "I ran out of strips… when I had strips and everything… I did. I don't do it no more. Because I know when it's high, and I know when it's low." (P12) • The Libre was what I was using. I'm not using anything right now. I had ran out… and I found out that the military doesn't cover that… that's why it's taking me so long to get it back. (P4)

[a]CGM: continuous glucose monitoring.

Behavioral Barriers: Managing Diet

The most frequently reported challenges were related to dietary management. Participants described diet as a major challenge in managing diabetes in their daily lives (n=10). Many individuals struggled to resist ingrained eating habits and sweet cravings. While some mentioned that sweets were "getting around me" despite their efforts to limit intake (P12), others just compensated for it with an additional dose of insulin instead of practicing abstinence (P9). Participant (P11) reported having adequate nutritional knowledge but lacking appropriate behavioral execution, citing situational limitations on healthy options and a lack of willpower in choosing better alternatives. For some, knowledge gaps in interpreting food labels and the fear of eating incorrectly were concerns. Participant (P8) reported that the initial diagnosis of diabetes brought overwhelming confusion regarding grocery shopping, reading labels, and food selection. Additional concerns about managing diet included situational limitations such as a busy work schedule (P5) and struggles with daily meal planning (P3), which led to choosing convenient food over meeting nutritional requirements.

Physical and Environmental Barriers: Mobility Constraints

Physical activity was hindered by physical limitations, including low energy, fatigue, limb pain, weakness, and visual impairment, as well as by environmental factors for some participants (n=4). Diabetes-related complications, such as leg pain and weakness, impaired mobility (P7), and ankle swelling, further hindered ambulation (P1) for participants. Extreme climates also interfered with outdoor walking routines for some participants (n=3), as participant P11 mentions: "I try to walk. I haven't been doing that lately. (Because it is cold outside)." Participants not limited by physical limitations reported engaging in some

XSL•FO
RenderX

mild-to-moderate physical activity. Walking emerges as the most common and accessible form of exercise.

Structural Barriers: Medication Access and Monitoring

Medication adherence was generally good among the participants in this study; however, structural barriers disrupted treatment access for some individuals. Barriers related to medication access included the unavailability of prescribed drugs and lack of insurance coverage (n=3). These challenges resulted in delays or changes in prescribed regimens. The available and potentially alternative medicines either had high copays, were not affordable for patients, or were not covered by insurance at all. Such experiences highlight financial barriers, insurance limitations, and medication shortages, leading to disturbances in adherence, even in highly motivated individuals. In these situations, participants often relied on health care providers for alternative solutions.

Glucose monitoring also emerged as an essential component of diabetes self-management for most participants, who incorporated blood glucose checks into their daily routines using traditional glucometers or continuous glucose monitoring (CGM) systems. However, structural barriers similarly affected access to monitoring tools (n=3). Challenges included supply shortages and insurance-related limitations on obtaining glucometers or CGM devices. Some participants expressed difficulty maintaining regular monitoring due to a lack of supplies or discomfort with fingerstick testing. Participant P12 discontinued fingerstick testing after exhausting test strips and now relies on symptom recognition for hyperglycemia or hypoglycemia. Participant P4 faced insurance barriers to receiving CGM monitors and mentioned that they are in the process of resolving access issues.

Theme Summary

Overall, the findings highlight that diabetes self-management is influenced by a complex interplay of behavioral, physical, and structural factors. Participants faced persistent difficulties in adhering to dietary recommendations, engaging in consistent physical activity, accessing and affording prescribed medications, and maintaining regular blood glucose monitoring. These challenges were compounded by long-standing behaviors, physical comorbidities, environmental constraints, and systemic barriers within health care and insurance systems. The convergence of these individual and contextual factors underscores the multifaceted nature of diabetes self-management and reveals the necessity for comprehensive, patient-centered interventions that address not only education and motivation but also accessibility, affordability, and continuity of care.

Theme 2. Future Interventions: Expectations Surrounding Telephone-Based DHC for Diabetes Management

Theme Overview

The multifaceted nature of diabetes self-management necessitated understanding patients' needs and expectations for an efficient, personalized telephone-based DHC program. Four subthemes emerged within this theme: (1) targeted education, (2) perception of a health coach, (3) delivery modality, and (4) incentive option. A brief description of each subtheme is provided in Table 3.

XSL•FO
RenderX

Table . Future interventions: expectations surrounding telephone-based digital health coaching for diabetes management.

Subthemes	Description	Quotes
Targeted education	• What to eat • Specific exercises • Day-to-day regime • Medication effect • All about diabetes • None • Guidance on specific topics such as meal planning, exercises, and medication side effects	• "I need to know more about my varieties, …I was in question about sweet potatoes and white potatoes. I didn't really get a good clarification whether I can eat those two or not, and different other things." (P3) • "I would like to know what specifically exercises that I can do if I ain't doing the right ones." (P7) • "I would like to know before they start... Okay. Say, for instance, they give you injections. Because before I was on injections, I was on pills. So I would like for the doctor or whomever to talk to you more about what the injections can do to your body or what it may make you feel like once you don't have it because I didn't start to feel like the drain and all of that until I wasn't able to take my meds." (P10)
Perception of a health coach	• Accountability partner to provide feedback and tailored meal preparation strategies	• "I learn best by hands-on.… if I'm talking to somebody that's telling me, "Okay, this is what I want you to do. You can actually print this out and this is your breakfast, this is your lunch, this is your dinner and snacks," and so on and so forth. And I can go shopping and get these things and actually pre-prepare my meals and things like that." (P4) • "I think it'll keep me... help me make better choices, keep me informed about what's out there, and what's available for me to control my diabetes." (P11) • "Of course. Well, just having someone to hold you accountable is always a good thing. Someone you can ask questions to about certain things would always benefit me." (P5)
Delivery modality	• Prefer text reminders with options for short, high-impact calls	• "Sometimes I forget, I have to ask my husband did he see me take my medicine because sometimes I forget and I'd be like, "Okay, I don't know if I remembered to take it. Did I take it or did I not?" So I think it [daily call/text] would help. That would be beneficiary helping me." (P9) • "Just to make sure you're doing it to stay on track, and make sure you're doing what you're supposed to do. I think it [weekly calls] would be vital. I think it'd be a help." (P11)
Incentive options	• Monetary rewards are mainly preferred, while some prefer gym memberships	• "… I think it would help me a lot financially. Because I'm on a fixed income, and I just got a prescription that I need to get filled. … The money would come in handy." (P8) • "Probably a membership to a gym or something [should be appropriate reward]." (P10) • "I'm not really worried about that [reward]. I'm more so worried about my health." (P3)

XSL•FO
RenderX

Targeted Education

Based on their experiences with current T2DM management strategies, the participants were asked whether they would like to incorporate specific knowledge into future T2DM interventions to help them monitor and manage their diabetes. Participants expressed targeted interests in diabetes education topics, with dietary guidance emerging as the most common need (n=3), particularly clarification on suitable foods for individuals with diabetes, as P7 mentions, “And then I can see about my eating habits, too, see what I can eat….what I can’t eat...I know a little of it, but I don’t know a lot of it.” Participants also emphasized the need to learn day-to-day regimens to build sustainable habits (n=2), noting the difficulty of changing long-standing lifestyles and the value of clear daily guidance for effective habit formation. Additional knowledge on other topics included specific exercises (n=1), side effects of diabetes medication (n=1), and the desire for overall refreshers (n=3). However, some indicated no need for additional information (n=4), citing self-learning, prior knowledge, or prior participation in research studies as sufficient.

Upon querying participants’ thoughts on receiving a customized booklet containing diabetes education information, accompanied by a follow-up call to explain its contents, most participants expressed interest in learning more and felt it would be a valuable tool (n=11). They mentioned that the booklet would serve as a quick, convenient guide for finding any information they needed compared to searching online on an electronic device.

Perceptions of a Health Coach

The concept of a health coach was introduced to the participants as a personal cheerleader, a guide, and a partner who could help them implement lifestyle changes to improve their health. When asked, the participants were asked whether they thought they would benefit from consulting with a health coach to manage diabetes. Most participants (n=8) were open to working with a health coach and expressed interest in tailored behavioral support. A few (n=2) were unsure of the idea and were willing to try if recommended, while only a couple (n=2) declined the idea of working with a coach, citing sufficient self-control and an aversion to added stress. Most participants perceived two primary benefits of working with a health coach: (1) an accountability partner for informed decision-making and keeping them on the right track and (2) providing support on personalized meal planning.

Participants also mentioned that receiving health coaching would help them develop strategies to manage their diabetes. They perceived that communicating information about their diabetes management to a health coach would help them receive constructive feedback, which, in turn, would enable them to manage their diabetes more effectively.

Delivery Modality: Use of Technology in Terms of Phone Calls or Text

Participants perceived using technology, such as a phone call or text, as part of the intervention would be vital.

Participants thought daily monitoring texts or phone calls could be helpful reminders because they had memory issues (n=10). They mentioned having difficulty remembering to take medications on time or checking their blood sugar levels. When asked about their preferred communication mode, participants preferred text over calls (n=7). Participants who favored texts did not like the idea of regular phone calls, as P9 quoted, “Probably a text because I don’t really like phone calls, so I’d rather text than take a phone call.” Participants who opted for calls (n=5) preferred quick interactions, as P10 stated, “It would be preferably a call because text messages, I’m not going to lie, they’re easy for me to ignore. So yeah.” Although most participants preferred receiving a daily text over a daily call, we also asked participants about the desired length of a daily call if they were to receive one; most specified that they wanted it to be short. The majority of participants preferred the call duration to be less than 10 minutes (<2 minutes [n=4], 5 - 10 minutes [n=4], as per the needs of the participants [n=1]).

Participants were also asked about their preference for weekly contact for coaching calls. Most participants (n=11) indicated that these calls would be a helpful component of the intervention, providing accountability and reminders to support routine building. One participant preferred a reduced frequency (once or twice a month), suggesting that weekly calls might be too repetitive and frustrating (P1).

Incentive Structure

Participants identified practical incentives to enhance engagement with daily calls or texts, with monetary rewards emerging as the predominant preference (n=10). Most favored financial support through cash or gift cards, with monetary rewards ranging from US $10 to US $50, viewing it as valuable supplemental income. P8 mentions: “...it would help me a lot financially... The money would come in handy.” A minority prioritized health over rewards (n=2), with P3 stating: “I’m not really worried about that [reward]. I’m more so worried about my health,” and P10 preferring alternatives that promote healthy behavior, such as a gym membership.

Theme Summary

Participants described their expectations surrounding a potential DHC intervention within this theme. Given the challenges they face in managing diabetes, they expected future interventions to incorporate specific information related to diet, day-to-day regimens, exercise, and medication effects to optimally manage their diabetes. Most participants identified several aspects of health coaching that would benefit them in managing diabetes. They perceived that a health coach could serve as an accountability partner, a guide for meal planning, and assistance in staying on track with diabetes management, while also providing a means to discuss aspects of diabetes beyond their PCP. Most participants found that using technology, such as daily or weekly calls and texts, would help them as reminders to take medications or exercise, except for a few who indicated that daily or weekly communications could be overwhelming. Most participants stated that a monetary incentive of US $10 to US $50 would motivate them to attend those calls, while a few noted that a financial incentive might not be as effective as other incentives, such as a gym membership.

XSL•FO
RenderX

Discussion

Principal Findings and Implications for Telephone-Based Digital Health Coaching

This qualitative study explored the lived experiences of individuals with T2DM, focusing on the daily challenges faced in achieving optimal glycemic control. It also examined participants' perceptions of a prospective health coaching intervention aimed at facilitating effective diabetes self-management and promoting long-term glycemic stability. The challenges identified among participants were consistent with findings from previous studies, highlighting the interplay of behavioral, physical, environmental, and structural factors that collectively impede effective diabetes self-management [43-45]. Insights and needs assessment findings from this study informed the conceptualization, design, and development of an AI-assisted telephone-based health-coaching intervention GODART pilot and feasibility trial (NCT05344859) to address the self-management gaps identified by participants [32]. Informed by participants' preferences for accessible, personalized support, the GODART intervention delivered daily IVR calls via an automated behavior-monitoring system to track diet, physical activity, medication adherence, and blood glucose monitoring, along with weekly coaching.

The persistent difficulties participants described with self-management of T2DM included adhering to dietary recommendations, integrating regular physical activity into daily routines, affording and accessing medications, and sustaining consistent blood glucose monitoring. These challenges were often rooted in long-standing eating patterns [46,47], limited nutritional literacy [48-51], physical limitations [52,53] and fatigue [54-56], environmental barriers to exercise [53,57,58], insurance-related constraints [59], and intermittent access to monitoring supplies or continuous glucose monitoring devices [60]. Taken together, the results underscore that diabetes self-management is not simply a matter of individual motivation but is embedded within broader socioeconomic, physical, and health system constraints that complicate efforts to maintain optimal glycemic control [61-64].

Participants expressed a strong interest in a future telephone-based DHC program aimed at providing targeted diabetes education and bridging the practical and informational gap. They articulated a desire for nuanced, personalized guidance on what to eat, how to interpret food labels, how to choose safe and realistic forms of physical activity given their functional limitations, and how diet, exercise, and medications interact to affect blood sugar. They also expressed a desire for clear guidance on building day-to-day routines that could help them overcome long-standing habits and develop sustainable, long-term lifestyle changes. This finding aligns with evidence highlighting the importance of personalized, practical information that addresses patients' immediate challenges and promotes adherence to self-management practices [65-67]. The emphasis on day-to-day routines suggests that participants value straightforward, actionable education over general advice [67-69]. Such person-centered approaches have been shown to improve comprehension, motivation, and glycemic control, especially when they align with individuals' health literacy levels and behavioral readiness [70-72].

We also found an overwhelmingly positive attitude toward the idea of receiving a customized diabetes education booklet. All participants unanimously viewed it as a valuable tool for learning more about diabetes, serving both as an overall refresher and a practical guide for managing diabetes [73,74]. This enthusiastic response highlights the importance of providing educational resources that are concise, visually engaging, and tailored to individuals' literacy levels [75]. Traditional diabetes education materials, which often contain dense or highly technical content, frequently fail to capture patients' attention or promote sustained learning [75-78]. In contrast, developing simplified, low-literacy educational packets may be more effective for improving understanding and facilitating long-term behavioral change among adults with type 2 diabetes [74,77,79]. In response, the GODART study developed a pictorial, low-literacy educational packet that incorporated materials from the Diabetes Literacy and Numeracy Education Toolkit [80] and the Centers for Disease Control and Prevention's National Diabetes Prevention Program [81] to ensure accessible, consistent foundational knowledge for all participants.

Our participants described a health coach as a nonjudgmental, accountable partner and a guide who helps improve health and implement lifestyle changes [82,83]. Most participants expressed a clear preference for working with a health coach to support behavior change, while a few were uncertain but willing to try one if recommended, suggesting that participants were generally open to modifying their behaviors when adequate support was available [19,84]. The effectiveness of health coaching programs appears greater when they are delivered more frequently within long-term interventions [20,85,86]. Frequent check-ins, typically held weekly or bi-weekly [87,88], enable timely adjustments to SMART (specific, measurable, attainable, realistic, and timely) goals based on participants' real-world challenges and help reduce overall stress levels [89-92]. Based on these responses, all participants in the GODART study received 24 weekly personalized health-coaching calls that guided them in setting progressive, achievable milestones for sustainable behavior change.

Regular behavior monitoring via texts, calls, or apps has been shown to enhance accountability and awareness in diabetes management for medication adherence, blood glucose monitoring, dietary behavior, and physical activity [93-96]. Echoing this evidence, most participants endorsed the idea that daily reminders, such as text messages or phone calls, may help them remember to take their medications on time and check their blood sugar as recommended. Only one participant expressed reluctance toward behavior monitoring via texts, calls, or apps due to concerns about frustration and repetition of the calls [97]. In response, we developed an automated behavior-monitoring system delivering daily IVR calls using AI-assisted natural language processing for the GODART study, where participants reported daily dietary intake, step counts, exercise duration, medication adherence, and glucose monitoring captured via natural speech, keypad inputs for numeric data, and simple yes or no responses [98-100]. To address call-related

XSL•FO
RenderX

frustrations, participants selected their preferred call times and had flexible return-call options, making it low burden while supporting consistent engagement [97].

Participants identified incentives as strong motivators for engaging with DHC texts or calls [101,102]. Preferences varied, with most participants reporting that monetary rewards were particularly motivating, while others emphasized nonmonetary incentives or intrinsic motivation related to health improvement. The diversity in motivational drivers is consistent with the existing literature, which shows that financial incentives can enhance short-term engagement and adherence to health interventions, particularly among individuals facing socioeconomic challenges [101,103]. Additionally, although both monetary and nonmonetary incentives were motivating, financial rewards appeared to exert a stronger influence than psychological motivators [104]. Based on this feedback, the GODART study incorporated financial incentives of up to US $100.80 to encourage adherence to daily monitoring calls over a 24-week intervention period.

Although participants' challenges with diet, physical activity, and structural barriers to diabetes self-management largely mirror existing qualitative literature, this formative study makes 2 key contributions. First, it focuses specifically on expectations for a low-tech, telephone-based DHC model intended to bridge the digital divide for adults with T2DM who have limited broadband access, smartphone use, or mobile literacy. Second, the study demonstrates how participant-identified needs, such as targeted diabetes education, health coaches as accountability partners, weekly coaching check-ins, low-literacy educational materials, daily reminders, and financial incentives, can be translated into concrete components of a telephone-based, AI-assisted IVR intervention. By systematically integrating and prioritizing patient perspectives over researcher-driven assumptions, this formative qualitative work directly informed the design of the GODART optimization and feasibility trial, illustrating how qualitative inquiry can be operationalized into a scalable, equity-oriented DHC intervention.

Limitations

The study's findings are based on a purposively selected sample from a single health system in the Deep South state of Alabama, which may limit the generalizability of the results to a broader population with T2DM in the Deep South. The sample size of this exploratory study was relatively small (n=12), and participant demographics were relatively homogeneous across age, race/ethnicity, socioeconomic status, and educational attainment. This homogeneity may have increased the risk of selection bias and limits the breadth of perspectives captured, potentially overlooking key topics, barriers, or preferences relevant to underrepresented subgroups—such as rural residents, diverse ethnicities, or individuals with varying levels of health literacy.

All participants were recruited from a single academic health system, the UAB, and were required to have access to a phone and be able to speak English to participate in the interview. This requirement potentially excluded individuals with limited access to health care or digital technology, as well as those from non–English-speaking backgrounds. Additionally, the study relied on self-reported data from semistructured interviews, which may be subject to recall and social desirability bias, as participants might have overreported or underreported certain behaviors or attitudes. Finally, while the study provides valuable insights into patient expectations for DHC, it does not assess the actual implementation or effectiveness of such interventions. Future research should include larger, more diverse samples and employ longitudinal or mixed methods approaches to better understand the evolving needs and outcomes of individuals with T2DM, particularly as digital health solutions continue to expand.

Conclusion

Our findings highlight the multifaceted daily challenges adults with T2DM face in achieving glycemic control. The study also reveals participants' strong interest in accessible, personalized, telephone-based health coaching interventions as a supportive tool for managing their condition through lifestyle counseling and psychosocial support. By translating participant insights into the design of the GODART pilot and feasibility trial, including low-literacy education, AI-assisted daily behavior monitoring, human vs automated coaching, and fixed vs adaptive incentives, this research exemplifies patient-centered optimization within the MOST (Multiphase Optimization Strategy) framework. These findings not only affirm the feasibility of scalable digital health solutions but also underscore the critical role of preparatory qualitative work in bridging self-management gaps to foster sustainable behavior change and long-term health equity. Future studies should build on this foundation to evaluate effectiveness across diverse settings and refine components for broader implementation.

Acknowledgments

The authors acknowledge the participation of all individuals involved in this study. The content presented herein is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of Diabetes and Digestive and Kidney Diseases. No AI tool was used at any stage in the preparation of the manuscript.

Funding

This research was supported by the National Institute of Diabetes and Digestive and Kidney Diseases under grant number R01DK129378. The funding organizations had no role in the study design, data collection and analysis, manuscript preparation, or the decision to submit the manuscript for publication.

XSL•FO
RenderX

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this study. Mehta has received consulting fees from The Obesity Society, Novo Nordisk, Heart Rhythm Clinical Research Solutions, and the New Balance Foundation Obesity Prevention Center.

Multimedia Appendix 1
Interview guide.
[DOCX File, 18 KB - diabetes_v11i1e81441_app1.docx]

References

1. National diabetes statistics report. Centers for Disease Control and Prevention (CDC). 2026. URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2026-07-14]
2. Davidson JA. The increasing role of primary care physicians in caring for patients with type 2 diabetes mellitus. Mayo Clin Proc 2010 Dec;85(12 Suppl):S3-S4. [doi: 10.4065/mcp.2010.0466] [Medline: 21106869]
3. Shrivastav M, Gibson W Jr, Shrivastav R, et al. Type 2 diabetes management in primary care: the role of retrospective, professional continuous glucose monitoring. Diabetes Spectr 2018 Aug;31(3):279-287. [doi: 10.2337/ds17-0024] [Medline: 30140145]
4. Naous E, Achkar A, Morin RA, Mitri J. A systematic review on the impact of electronic consultations on diabetes care. Endocr Pract 2025 Jun;31(6):829-838. [doi: 10.1016/j.eprac.2025.03.001] [Medline: 40057305]
5. Kumar R, Kumar Gupta P, Gupta A, Randenikumara S. Role of primary care physicians in the prevention and control of liver diseases. In: Bhadoria AS, Nundy S, editors. Preventive Hepatology: CRC Press; 2025:194-204.
6. Zeytinoglu M, Huang ES. Diabetes: a primary health care approach. In: Andrade J, Pinto F, Arnett D, editors. Prevention of Cardiovascular Diseases: From Current Evidence to Clinical Practice: Springer; 2015:91-99. [doi: 10.1007/978-3-319-22357-5_10]
7. Tai-Seale M, McGuire TG, Zhang W. Time allocation in primary care office visits. Health Serv Res 2007 Oct;42(5):1871-1894. [doi: 10.1111/j.1475-6773.2006.00689.x] [Medline: 17850524]
8. Espinoza P, Varela CA, Vargas IE, et al. The burden of treatment in people living with type 2 diabetes: a qualitative study of patients and their primary care clinicians. PLoS One 2020;15(10):e0241485. [doi: 10.1371/journal.pone.0241485] [Medline: 33125426]
9. Nguyen MLT, Honcharov V, Ballard D, Satterwhite S, McDermott AM, Sarkar U. Primary care physicians' experiences with and adaptations to time constraints. JAMA Netw Open 2024 Apr 1;7(4):e248827. [doi: 10.1001/jamanetworkopen.2024.8827] [Medline: 38687477]
10. Rushforth B, McCrorie C, Glidewell L, Midgley E, Foy R. Barriers to effective management of type 2 diabetes in primary care: qualitative systematic review. Br J Gen Pract 2016 Feb;66(643):e114-e127. [doi: 10.3399/bjgp16X683509] [Medline: 26823263]
11. Koenigsberg MR, Corliss J. Diabetes self-management: facilitating lifestyle change. Am Fam Physician 2017 Sep 15;96(6):362-370. [Medline: 28925635]
12. Fuller H, Alberti H. Barriers to lifestyle changes in people with diabetes. Br J Gen Pract 2017 Feb;67(655):61. [doi: 10.3399/bjgp17X689005] [Medline: 28126861]
13. Kalra S, Jena BN, Yeravdekar R. Emotional and psychological needs of people with diabetes. Indian J Endocrinol Metab 2018;22(5):696-704. [doi: 10.4103/ijem.IJEM_579_17] [Medline: 30294583]
14. Bugg II KC. Cognitive-behavioral therapy manual to address glucose levels in aging diabetics: a pilot study [Dissertation]. : Fuller Theological Seminary, School of Psychology; 2005 URL: https://www.proquest.com/openview/680b064a3b58ad369d3dce38388f293e/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2026-07-14]
15. Beverly EA, Wray LA, Chiu CJ, Weinger K. Perceived challenges and priorities in co-morbidity management of older patients with type 2 diabetes. Diabet Med 2011 Jul;28(7):781-784. [doi: 10.1111/j.1464-5491.2011.03282.x] [Medline: 21395674]
16. Pantalone KM, Hobbs TM, Wells BJ, et al. Clinical characteristics, complications, comorbidities and treatment patterns among patients with type 2 diabetes mellitus in a large integrated health system. BMJ Open Diabetes Res Care 2015;3(1):e000093. [doi: 10.1136/bmjdrc-2015-000093] [Medline: 26217493]
17. Morley JE, Abbatecola AM, Woo J. Management of comorbidities in older persons with type 2 diabetes. J Am Med Dir Assoc 2017 Aug 1;18(8):639-645. [doi: 10.1016/j.jamda.2017.05.006] [Medline: 28625388]
18. Hood KK, Hilliard M, Piatt G, Ievers-Landis CE. Effective strategies for encouraging behavior change in people with diabetes. Diabetes Manag (Lond) 2015;5(6):499-510. [Medline: 30100925]
19. Matthews JA, Matthews S, Faries MD, Wolever RQ. Supporting sustainable health behavior change: the whole is greater than the sum of its parts. Mayo Clin Proc Innov Qual Outcomes 2024 Jun;8(3):263-275. [doi: 10.1016/j.mayocpiqo.2023.10.002] [Medline: 38807973]

XSL•FO
RenderX

20. Racey M, Jovkovic M, Alliston P, Ali MU, Sherifali D. Diabetes health coach in individuals with type 2 diabetes: a systematic review and meta analysis of quadruple aim outcomes. Front Endocrinol (Lausanne) 2022;13:1069401. [doi: 10.3389/fendo.2022.1069401] [Medline: 36589795]
21. Mobile fact sheet. Pew Research Center. 2025. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/?cb_viewport=tablet [accessed 2026-07-14]
22. Lee DY, Park J, Choi D, Ahn HY, Park SW, Park CY. The effectiveness, reproducibility, and durability of tailored mobile coaching on diabetes management in policyholders: a randomized, controlled, open-label study. Sci Rep 2018 Feb 26;8(1):3642. [doi: 10.1038/s41598-018-22034-0] [Medline: 29483559]
23. He Q, Zhao X, Wang Y, Xie Q, Cheng L. Effectiveness of smartphone application-based self-management interventions in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2022 Feb;78(2):348-362. [doi: 10.1111/jan.14993] [Medline: 34324218]
24. Hyun MK, Lee JW, Ko SH, Hwang JS. Improving glycemic control in type 2 diabetes using mobile applications and e-coaching: a mixed treatment comparison network meta-analysis. J Diabetes Sci Technol 2022 Sep;16(5):1239-1252. [doi: 10.1177/19322968211010153] [Medline: 33980055]
25. Gerber BS, Biggers A, Tilton JJ, et al. Mobile health intervention in patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023 Sep 5;6(9):e2333629. [doi: 10.1001/jamanetworkopen.2023.33629] [Medline: 37773498]
26. Hsia J, Guthrie NL, Lupinacci P, et al. Randomized, controlled trial of a digital behavioral therapeutic application to improve glycemic control in adults with type 2 diabetes. Diabetes Care 2022 Dec 1;45(12):2976-2981. [doi: 10.2337/dc22-1099] [Medline: 36181554]
27. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally [accessed 2026-07-14]
28. Hong YA, Cho J. Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]
29. Kumar D, Hemmige V, Kallen MA, Giordano TP, Arya M. Mobile phones may not bridge the digital divide: a look at mobile phone literacy in an underserved patient population. Cureus 2019 Feb 20;11(2):e4104. [doi: 10.7759/cureus.4104] [Medline: 31057998]
30. Gazmararian JA, Yang B, Elon L, Graham M, Parker R. Successful enrollment in Text4Baby more likely with higher health literacy. J Health Commun 2012;17 Suppl 3:303-311. [doi: 10.1080/10810730.2012.712618] [Medline: 23030578]
31. Poorman E, Gazmararian J, Elon L, Parker R. Is health literacy related to health behaviors and cell phone usage patterns among the text4baby target population? Arch Public Health 2014;72(1):13. [doi: 10.1186/2049-3258-72-13] [Medline: 24872883]
32. Mehta T, John T, El Zein A, et al. Gamified Optimized Diabetes Management with Artificial Intelligence-Powered Rural Telehealth Intervention (GODART): protocol for an optimization pilot and feasibility trial. JMIR Res Protoc 2025 Dec 5;14(1):e70271. [doi: 10.2196/70271] [Medline: 41348456]
33. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th edition: Sage publications; 2018.
34. rev. URL: https://www.rev.com/ [accessed 2026-07-20]
35. Bartholomew TT, Joy EE, Kang E, Brown J. A choir or cacophony? Sample sizes and quality of conveying participants' voices in phenomenological research. Methodol Innov 2021 May;14(2):20597991211040063. [doi: 10.1177/20597991211040063]
36. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis: SAGE Publications, Inc; 2012. [doi: 10.4135/9781483384436]
37. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
38. Merriam SB, Grenier RS. Qualitative Research in Practice: Examples for Discussion and Analysis, 2nd edition: John Wiley & Sons; 2019.
39. Ridder HG. Book review: qualitative data analysis. A methods sourcebook. Ger J Hum Resour Manag 2014 Nov;28(4):485-487. [doi: 10.1177/239700221402800402]
40. Boeije HR. Analysis in Qualitative Research: Sage Publications; 2009.
41. El Zein A, Garla V, Hall ME, et al. Rationale and design of the FREEDOM study: a hybrid type 1 optimization-implementation trial to improve type 2 diabetes management in primary care. Contemp Clin Trials 2026 Feb;161:108173. [doi: 10.1016/j.cct.2025.108173] [Medline: 41354173]
42. El Zein A, Hays D, Glidden G, et al. A CFIR-guided qualitative study of digital health engagement among Black adults with type 2 diabetes. NPJ Digit Med 2025 Dec 10;8(1):755. [doi: 10.1038/s41746-025-02107-x] [Medline: 41372595]

XSL•FO
RenderX

43. Masupe T, Onagbiye S, Puoane T, Pilvikki A, Alvesson HM, Delobelle P. Diabetes self-management: a qualitative study on challenges and solutions from the perspective of South African patients and health care providers. Glob Health Action 2022 Dec 31;15(1):2090098. [doi: 10.1080/16549716.2022.2090098] [Medline: 35856773]
44. Ngo AD, Kulintang MB, Alharbi TAF, et al. Challenges and barriers in managing type 2 diabetes: insights from a qualitative study of patients' experiences at a university hospital in Qassim, Saudi Arabia. Glob Health Action 2026 Dec;19(1):2591543. [doi: 10.1080/16549716.2025.2591543] [Medline: 41504720]
45. Woodward A, Walters K, Davies N, et al. Barriers and facilitators of self-management of diabetes amongst people experiencing socioeconomic deprivation: a systematic review and qualitative synthesis. Health Expect 2024 Jun;27(3):e14070. [doi: 10.1111/hex.14070] [Medline: 38751247]
46. Mostafavi-Darani F, Zamani-Alavijeh F, Mahaki B, Salahshouri A. Exploring the barriers of adherence to dietary recommendations among patients with type 2 diabetes: a qualitative study in Iran. Nurs Open 2020 Nov;7(6):1735-1745. [doi: 10.1002/nop2.558] [Medline: 33072357]
47. Vijan S, Stuart NS, Fitzgerald JT, et al. Barriers to following dietary recommendations in type 2 diabetes. Diabet Med 2005 Jan;22(1):32-38. [doi: 10.1111/j.1464-5491.2004.01342.x] [Medline: 15606688]
48. Mahmoodi MR, Khanjani N. Barriers and limitations to obstacle diabetes self-management with a focus on nutritional literacy: solutions and opportunities. Critical review and research synthesis. Crit Comments Biomed 2020;1(1):e10008. [doi: 10.18502/ccb.v1i1.2870]
49. Hashim SA, Barakatun-Nisak MY, Abu Saad H, Ismail S, Hamdy O, Mansour AA. Association of health literacy and nutritional status assessment with glycemic control in adults with type 2 diabetes mellitus. Nutrients 2020 Oct 15;12(10):3152. [doi: 10.3390/nu12103152] [Medline: 33076406]
50. Caruso R, Magon A, Baroni I, et al. Health literacy in type 2 diabetes patients: a systematic review of systematic reviews. Acta Diabetol 2018 Jan;55(1):1-12. [doi: 10.1007/s00592-017-1071-1] [Medline: 29129000]
51. Siopis G, Colagiuri S, Allman-Farinelli M. People with type 2 diabetes report dietitians, social support, and health literacy facilitate their dietary change. J Nutr Educ Behav 2021 Jan;53(1):43-53. [doi: 10.1016/j.jneb.2020.09.003] [Medline: 33077370]
52. Ahmad E, Sargeant JA, Yates T, Webb DR, Davies MJ. Type 2 diabetes and impaired physical function: a growing problem. Diabetology 2022;3(1):30-45. [doi: 10.3390/diabetology3010003]
53. Lidegaard LP, Schwennesen N, Willaing I, Faerch K. Barriers to and motivators for physical activity among people with type 2 diabetes: patients' perspectives. Diabet Med 2016 Dec;33(12):1677-1685. [doi: 10.1111/dme.13167] [Medline: 27279343]
54. Singh R, Teel C, Sabus C, McGinnis P, Kluding P. Fatigue in type 2 diabetes: impact on quality of life and predictors. PLoS One 2016;11(11):e0165652. [doi: 10.1371/journal.pone.0165652] [Medline: 27824886]
55. Fritschi C, Quinn L, Hacker ED, et al. Fatigue in women with type 2 diabetes. Diabetes Educ 2012;38(5):662-672. [doi: 10.1177/0145721712450925] [Medline: 22713262]
56. Ba J, Chen Y, Liu D. Fatigue in adults with type 2 diabetes: a systematic review and meta-analysis. West J Nurs Res 2021 Feb;43(2):172-181. [doi: 10.1177/0193945920938636] [Medline: 32627705]
57. Korkiakangas EE, Alahuhta MA, Laitinen JH. Barriers to regular exercise among adults at high risk or diagnosed with type 2 diabetes: a systematic review. Health Promot Int 2009 Dec;24(4):416-427. [doi: 10.1093/heapro/dap031] [Medline: 19793763]
58. Thielen SC, Reusch JEB, Regensteiner JG. A narrative review of exercise participation among adults with prediabetes or type 2 diabetes: barriers and solutions. Front Clin Diabetes Healthc 2023;4:1218692. [doi: 10.3389/fcdhc.2023.1218692] [Medline: 37711232]
59. Goetjes E, Blankart KE. Insurance barriers and inequalities in health care access: evidence from dual practice. Health Econ Rev 2024 Mar 21;14(1):23. [doi: 10.1186/s13561-024-00500-y] [Medline: 38512590]
60. Bowker SL, Mitchell CG, Majumdar SR, Toth EL, Johnson JA. Lack of insurance coverage for testing supplies is associated with poorer glycemic control in patients with type 2 diabetes. CMAJ 2004 Jul 6;171(1):39-43. [doi: 10.1503/cmaj.1031830] [Medline: 15238494]
61. Gonzalez-Zacarias AA, Mavarez-Martinez A, Arias-Morales CE, Stoicea N, Rogers B. Impact of demographic, socioeconomic, and psychological factors on glycemic self-management in adults with type 2 diabetes mellitus. Front Public Health 2016;4:195. [doi: 10.3389/fpubh.2016.00195] [Medline: 27672634]
62. Stotz SA, Ricks KA, Eisenstat SA, Wexler DJ, Berkowitz SA. Opportunities for interventions that address socioeconomic barriers to type 2 diabetes management: patient perspectives. Sci Diabetes Self Manag Care 2021 Apr;47(2):153-163. [doi: 10.1177/0145721721996291] [Medline: 34078177]
63. Adu MD, Malabu UH, Malau-Aduli AEO, Malau-Aduli BS. Enablers and barriers to effective diabetes self-management: a multi-national investigation. PLoS One 2019;14(6):e0217771. [doi: 10.1371/journal.pone.0217771] [Medline: 31166971]
64. Tuobenyiere J, Mensah GP, Korsah KA. Patient perspective on barriers in type 2 diabetes self-management: a qualitative study. Nurs Open 2023 Oct;10(10):7003-7013. [doi: 10.1002/nop2.1956] [Medline: 37488987]
65. Alum EU. Optimizing patient education for sustainable self-management in type 2 diabetes. Discov Public Health 2025;22(1):44. [doi: 10.1186/s12982-025-00445-5]

XSL•FO
RenderX

66. Williams DM, Jones H, Stephens JW. Personalized type 2 diabetes management: an update on recent advances and recommendations. Diabetes Metab Syndr Obes 2022;15:281-295. [doi: 10.2147/DMSO.S331654] [Medline: 35153495]
67. Nguyen TNP, Thi CN, Phuong TPT, Ngoc QN, Pham HT. Counseling preferences among patients with type 2 diabetes: implications for personalized care. J Diabetes Res 2025;2025(1):9970845. [doi: 10.1155/jdr/9970845] [Medline: 40771771]
68. van Smoorenburg AN, Hertroijs DFL, Dekkers T, Elissen AMJ, Melles M. Patients' perspective on self-management: type 2 diabetes in daily life. BMC Health Serv Res 2019 Aug 28;19(1):605. [doi: 10.1186/s12913-019-4384-7] [Medline: 31462220]
69. Asmat K, Dhamani K, Gul R, Froelicher ES. The effectiveness of patient-centered care vs. usual care in type 2 diabetes self-management: a systematic review and meta-analysis. Front Public Health 2022;10:994766. [doi: 10.3389/fpubh.2022.994766] [Medline: 36388341]
70. Chiou SJ, Chang YJ, Chen CD, Liao K, Tseng TS. Using patient profiles for sustained diabetes management among people with type 2 diabetes. Prev Chronic Dis 2023 Mar 16;20:E13. [doi: 10.5888/pcd20.220210] [Medline: 36927708]
71. Swanson V, Maltinsky W. Motivational and behaviour change approaches for improving diabetes management. Pract Diab 2019 Jul;36(4):121-125. [doi: 10.1002/pdi.2229]
72. Prueksaritanond S, Tubtimtes S, Asavanich K, Tiewtranon V. Type 2 diabetic patient-centered care. J Med Assoc Thai 2004 Apr;87(4):345-352. [Medline: 15217167]
73. O'Donnell M, Alvarez-Iglesias A, McGuire BE, Dinneen SF. The impact of sharing personalised clinical information with people with type 2 diabetes prior to their consultation: a pilot randomised controlled trial. Patient Educ Couns 2016 Apr;99(4):591-599. [doi: 10.1016/j.pec.2015.11.014] [Medline: 26654869]
74. Saritessa N, Ibrahim K, Haroen H. Booklet-based education in diabetes mellitus patients: a scoping review. J Health Nutr Res 2025;4(1):172-185. [doi: 10.56303/jhnresearch.v4i1.372]
75. Yasin NSM, Mahmood TMT, Hatah E, Hazaidi HA, Abidin N. Tailoring educational content for T2DM patients: a qualitative study on preferences for interactive multimedia applications. J ASEAN Fed Endocr Soc 2025 May;40(1):45-52. [doi: 10.15605/jafes.040.01.18] [Medline: 40416481]
76. Reid JC, Klachko DM, Kardash CA, Robinson RD, Scholes R, Howard D. Why people don't learn from diabetes literature: influence of text and reader characteristics. Patient Educ Couns 1995 Feb;25(1):31-38. [doi: 10.1016/0738-3991(94)00688-i] [Medline: 7603931]
77. Smith KM, Baker KM, Bardsley JK, McCartney P, Magee M. Redesigning hospital diabetes education: a qualitative evaluation with nursing teams. J Nurs Care Qual 2019;34(2):151-157. [doi: 10.1097/NCQ.0000000000000349] [Medline: 30028413]
78. Heshmati H, Shakibazadeh E, Foroushani AR, Sadeghi R. A comprehensive model of health education barriers of health-care system in Iran. J Educ Health Promot 2020;9(1):106. [doi: 10.4103/jehp.jehp_23_20] [Medline: 32642462]
79. Wolff K, Cavanaugh K, Malone R, et al. The Diabetes Literacy and Numeracy Education Toolkit (DLNET): materials to facilitate diabetes education and management in patients with low literacy and numeracy skills. Diabetes Educ 2009;35(2):233-236. [doi: 10.1177/0145721709331945] [Medline: 19240246]
80. Cavanaugh K, Huizinga MM, Wolff K, et al. Diabetes Literacy and Numeracy Education Toolkit (DLNET): a randomized controlled trial of numeracy-focused enhanced diabetes education. Presented at: 68th Annual Meeting of the American-Diabetes-Association; Jun 6-10, 2008 URL: https://www.researchgate.net/publication/295195225_Diabetes_Literacy_and_Numeracy_Education_Toolkit_DLNET_A_randomized_controlled_trial_of_numeracy-focused_enhanced_diabetes_education [accessed 2026-07-14]
81. Ely EK, Gruss SM, Luman ET, et al. A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. Diabetes Care 2017 Oct;40(10):1331-1341. [doi: 10.2337/dc16-2099] [Medline: 28500215]
82. Perlman AI, Abu Dabrh AM. Health and wellness coaching in serving the needs of today's patients: a primer for healthcare professionals. Glob Adv Health Med 2020;9:2164956120959274. [doi: 10.1177/2164956120959274] [Medline: 33014630]
83. Liddy C, Johnston S, Irving H, Nash K, Ward N. Improving awareness, accountability, and access through health coaching: qualitative study of patients' perspectives. Can Fam Physician 2015 Mar;61(3):e158-e164. [Medline: 25932483]
84. DeJesus RS, Clark MM, Rutten LJF, et al. Wellness coaching to improve lifestyle behaviors among adults with prediabetes: patients' experience and perceptions to participation. J Patient Exp 2018 Dec;5(4):314-319. [doi: 10.1177/2374373518769118] [Medline: 30574554]
85. Sherifali D, Viscardi V, Bai JW, Ali RMU. Evaluating the effect of a diabetes health coach in individuals with type 2 diabetes. Can J Diabetes 2016 Feb;40(1):84-94. [doi: 10.1016/j.jcjd.2015.10.006] [Medline: 26827684]
86. Sherifali D, Bai JW, Kenny M, Warren R, Ali MU. Diabetes self-management programmes in older adults: a systematic review and meta-analysis. Diabet Med 2015 Nov;32(11):1404-1414. [doi: 10.1111/dme.12780] [Medline: 25865179]
87. Sforzo GA, Kaye MP, Faber A, Moore M. Dosing of health and wellness coaching for obesity and type 2 diabetes: research synthesis to derive recommendations. Am J Lifestyle Med 2023;17(3):374-385. [doi: 10.1177/15598276211073078] [Medline: 37304739]

XSL•FO
RenderX

88. Hohberg V, Kreppke JN, Kohl J, et al. Effectiveness of a personal health coaching intervention (diabetescoach) in patients with type 2 diabetes: protocol for an open-label, pragmatic randomised controlled trial. BMJ Open 2022 Jun 1;12(6):e057948. [doi: 10.1136/bmjopen-2021-057948] [Medline: 35649615]
89. Bahrami Z, Heidari A, Cranney J. Applying SMART goal intervention leads to greater goal attainment, need satisfaction and positive affect. Int J Ment Health Promot 2022;24(6):869-882. [doi: 10.32604/ijmhp.2022.018954]
90. Nguyen-Vaselaar H, Lappe JM. Goal-setting behavior for physical activity in adults with diabetes: a pilot project. J Nurse Pract 2021 Nov;17(10):1281-1284. [doi: 10.1016/j.nurpra.2021.08.015]
91. Cook HE, Garris LA, Gulum AH, Steber CJ. Impact of SMART goals on diabetes management in a pharmacist-led telehealth clinic. J Pharm Pract 2024 Feb;37(1):54-59. [doi: 10.1177/08971900221125021] [Medline: 36052449]
92. Misra R, Shawley-Brzoska S, Khan R, Kirk BO, Wen S, Sambamoorthi U. Addressing diabetes distress in self-management programs: results of a randomized feasibility study. J Appalach Health 2021;3(3):68-85. [doi: 10.13023/jah.0303.06] [Medline: 35770030]
93. Seiglie J, Tobolsky S, Crespo Trevino R, et al. Adapting a text messaging intervention to improve diabetes medication adherence in a Spanish-speaking population: qualitative study. JMIR Hum Factors 2025 May 1;12:e66668. [doi: 10.2196/66668] [Medline: 40311126]
94. Dobson R, Carter K, Cutfield R, et al. Diabetes text-message self-management support program (SMS4BG): a pilot study. JMIR mHealth uHealth 2015 Mar 25;3(1):e32. [doi: 10.2196/mhealth.3988] [Medline: 25830952]
95. Vervloet M, van Dijk L, Santen-Reestman J, et al. SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored. Int J Med Inform 2012 Sep;81(9):594-604. [doi: 10.1016/j.ijmedinf.2012.05.005] [Medline: 22652012]
96. Dobson R, Whittaker R, Jiang Y, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. BMJ 2018 May 17;361:k1959. [doi: 10.1136/bmj.k1959] [Medline: 29773539]
97. Lauffenburger JC, Barlev RA, Sears ES, et al. Preferences for mHealth technology and text messaging communication in patients with type 2 diabetes: qualitative interview study. J Med Internet Res 2021 Jun 11;23(6):e25958. [doi: 10.2196/25958] [Medline: 34114964]
98. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. Nurs Outlook 2003;51(6):277-283. [doi: 10.1016/s0029-6554(03)00161-1] [Medline: 14688763]
99. Yadav N, Rajpoot DS, Dhakad SK. LARAVEL: a PHP framework for e-commerce website. Presented at: 2019 Fifth International Conference on Image Information Processing (ICIIP); Nov 15-17, 2019. [doi: 10.1109/ICIIP47207.2019.8985771]
100. Natural language nutrition tracking demo. Mohanraj Thirumalai YouTube page. 2019 Dec 11. URL: https://www.youtube.com/watch?v=mz8nBDtpHho [accessed 2026-07-14]
101. Wurst R, Maliezefski A, Ramsenthaler C, Brame J, Fuchs R. Effects of incentives on adherence to a web-based intervention promoting physical activity: naturalistic study. J Med Internet Res 2020 Jul 30;22(7):e18338. [doi: 10.2196/18338] [Medline: 32729835]
102. Hulbert LR, Michael SL, Charter-Harris J, Atkins C, Skeete RA, Cannon MJ. Effectiveness of incentives for improving diabetes-related health indicators in chronic disease lifestyle modification programs: a systematic review and meta-analysis. Prev Chronic Dis 2022 Oct 27;19:E66. [doi: 10.5888/pcd19.220151] [Medline: 36302383]
103. Reid N, Nisenbaum R, Hwang SW, et al. The impact of financial incentives on service engagement among adults experiencing homelessness and mental illness: a pragmatic trial protocol. Front Psychiatry 2021;12:722485. [doi: 10.3389/fpsyt.2021.722485] [Medline: 34413804]
104. Medvedev D, Davenport D, Talhelm T, Li Y. The motivating effect of monetary over psychological incentives is stronger in WEIRD cultures. Nat Hum Behav 2024 Mar;8(3):456-470. [doi: 10.1038/s41562-023-01769-5] [Medline: 38191844]

Abbreviations

ADI: Area Deprivation Index
CGM: continuous glucose monitoring
DHC: digital health coaching
GODART: Gamified Optimized Diabetes Management with Artificial Intelligence–Powered Rural Telehealth
IVR: interactive voice response
MOST: Multiphase Optimization Strategy
NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases
PCP: primary care provider
SMART: specific, measurable, attainable, realistic, and timely
T2DM: type 2 diabetes mellitus
UAB: University of Alabama at Birmingham

XSL•FO
RenderX

Edited by R Zhao; submitted 04.Aug.2025; peer-reviewed by G Martin-Alemany, I Milani; revised version received 02.Jul.2026; accepted 03.Jul.2026; published 07.Aug.2026.

Please cite as:
John T, Mohanraj S, Nawshin T, Mehta T, Thirumalai M
Bridging the Digital Divide in Diabetes Care With an AI-Powered Health Coaching Intervention: Qualitative Formative Study
JMIR Diabetes 2026;11:e81441
URL: https://diabetes.jmir.org/2026/1/e81441
doi:10.2196/81441

© Tejossy John, Sangeetha Mohanraj, Tanjila Nawshin, Tapan Mehta, Mohanraj Thirumalai. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 7.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

GraphRAG-Enabled Local Large Language Model for Gestational Diabetes Mellitus: Development of a Proof-of-Concept

Edmund Evangelista[1*], PhD; Fathima Ruba[2*], MSc; Salman Bukhari[3*], BEng; Amril Nazir[1*], PhD; Ravishankar Sharma[1*], PhD

[1]College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates
[2]Department of Computer Science and Artificial Intelligence, Plaksha University, Chandigarh, India
[3]Department of Electrical Engineering, Capital University of Science and Technology, Islamabad, Pakistan
[*]all authors contributed equally

Corresponding Author:
Ravishankar Sharma, PhD
College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates

Abstract

Background: Gestational diabetes mellitus (GDM) is a prevalent chronic condition that affects maternal and fetal health outcomes worldwide, increasingly in underserved populations. While generative artificial intelligence (AI) and large language models (LLMs) have shown promise in health care, their application in GDM management remains underexplored.

Objective: This study aimed to investigate whether retrieval-augmented generation techniques, when combined with knowledge graphs (KGs), could improve the contextual relevance and accuracy of AI-driven clinical decision support. For this, we developed and validated a graph-based retrieval-augmented generation (GraphRAG)–enabled local LLM as a clinical support tool for GDM management, assessing its performance against open-source LLM tools.

Methods: A prototype clinical AI assistant was developed using a GraphRAG constructed from 1212 peer-reviewed research articles on GDM interventions, retrieved from the Semantic Scholar API (2000 - 2024). The GraphRAG prototype integrated entity extraction, KG construction using Neo4j, and retrieval-augmented response generation. The performance was evaluated in a simulated environment using clinical and layperson prompts, comparing the outputs of the systems against ChatGPT (OpenAI), Claude (Anthropic), and BioMistral models across 5 common natural language generation metrics.

Results: The GraphRAG-enabled local LLM showed higher accuracy in generating clinically relevant responses. It achieved a bilingual evaluation understudy score of 0.99, Jaccard similarity of 0.98, and BERTScore of 0.98, outperforming the benchmark LLMs. The prototype also produced accurate, evidence-based recommendations for clinicians and patients, demonstrating its feasibility as a clinical support tool.

Conclusions: GraphRAG-enabled local LLMs show much potential for improving personalized GDM care by integrating domain-specific evidence and contextual retrieval. Our prototype proof-of-concept serves two purposes: (1) the local LLM architecture gives practitioners from underserved locations access to state-of-the-art medical research in the treatment of chronic conditions and (2) the KG schema may be feasibly built on peer-reviewed, indexed publications, devoid of hallucinations and contextualized with patient data. We conclude that advanced AI techniques such as KGs, retrieval-augmented generation, and local LLMs improve GDM management decisions and other similar conditions and advance equitable health care delivery in resource-constrained health care environments.

(JMIR Diabetes 2026;11:e76454) doi:10.2196/76454

KEYWORDS

artificial intelligence for health care; generative AI; knowledge graph; retrieval augmented generation; large language model; gestational diabetes mellitus; explainable AI in medicine; GDM; artificial intelligence

Introduction

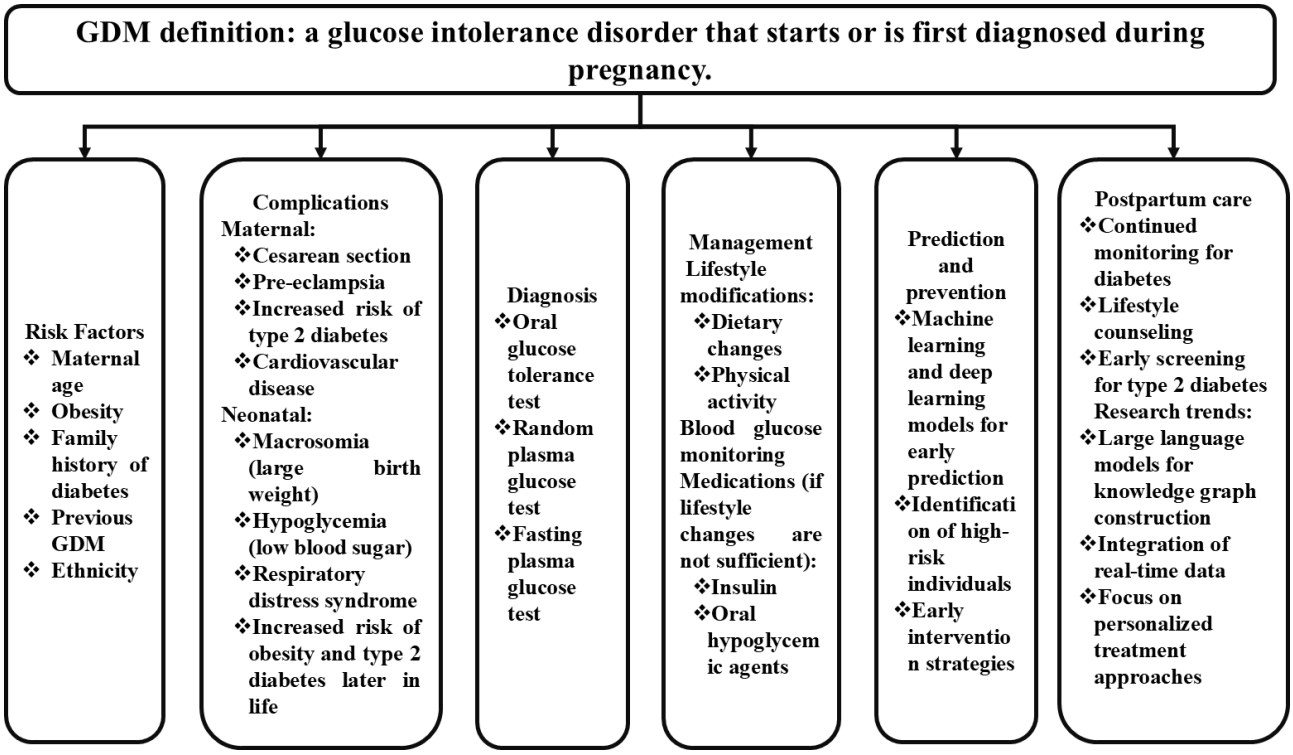
The growing use of electronic medical records linking diverse patient characteristics and prescription choices with positive treatment outcomes in large-scale use cases has resulted in platforms that guide optimal treatment options. For example, Sharma et al [1] presented an approach for delivering personalized health care as a means of effectively using scarce medical resources in underserved regions and populations, supporting the value of artificial intelligence (AI)–driven systems in such settings. While machine learning (ML) and data analytics have generated individualized treatment recommendations for improving outcomes, “these works focused on making broad [largely drug class level] treatment

XSL•FO
RenderX

recommendations independently of specific drug and dose considerations... [whereas] guidelines and landmark trials highlight important drug- and dose-dependent variations in treatment efficacy, safety, and risk profiles" [2]. In short, personalized medicine should account for contextual variations in seeking more effective, cost-efficient treatments with better outcomes. This study presents an approach to clinical support to time- and resource-constrained practitioners using a generative artificial intelligence (GenAI) approach to treat a serious medical condition afflicting young mothers and their children with increasing alacrity. Such a need is particularly acute in the socioeconomically disadvantaged regions of the world.

Gestational diabetes mellitus (GDM) is a significant global health concern affecting many pregnancies [3]. Defined as glucose metabolism imbalance first detected during pregnancy, the International Association of Diabetes in Pregnancy Study Group reports that "GDM is not only related to perinatal morbidity but also to an increased risk of diabetes and cardiovascular disease in the mother in later life, and childhood obesity in the offspring" [4]. The pooled global prevalence was 14% in 2021, with the highest occurrence in the Middle East - North Africa (27.6%), Southeast Asia (20.8%), and among high-income countries (14.2%) [5]. There is considerable agreement among medical practitioners that the development of GDM could be influenced by various risk factors, including maternal age, obesity, family history of diabetes, previous occurrences of GDM, and specific ethnic backgrounds [6,7]. This is illustrated in Figure 1 (data sources: [3,8-10]) as the medical characterizations of GDM comprising factors such as diagnosis, risks, prediction, management, complications, and postpartum care.

Figure 1. Medical characterization of gestational diabetes mellitus. GDM: gestational diabetes mellitus.



Also, of concern to the WHO is that GDM leads to various complications for both affected mothers and their offspring, such as increased risks of cesarean delivery, pre-eclampsia, and type 2 diabetes (T2D) for mothers. Children are at higher risk of macrosomia, hypoglycemia, respiratory distress syndrome, and an increased likelihood of developing obesity and T2D later in life [11]. The long-term health risks include elevated chances of developing T2D and cardiovascular diseases for both mother and child [12]. In the Global South and developing countries [8,13], GDM presents significant challenges due to:

1. Higher prevalence rates in certain regions, particularly South Asia and the Middle East.
2. Limited health care resources for screening, diagnosis, and management.
3. Genetic factors in certain ethnic groups increase GDM risk.
4. Rapid urbanization and lifestyle changes leading to increased obesity rates.
5. Potential underdiagnoses due to lack of routine screening.

Effective GDM treatment requires multiple diagnostic tests, including oral glucose tolerance tests, random plasma glucose tests, and fasting plasma glucose tests. The treatment options include regular blood glucose monitoring, dietary modifications, lifestyle changes, and, when necessary, pharmacological interventions such as insulin or oral hypoglycemic agents [9]. The recent advancements in AI-driven tools, such as the AI Drug Mix and Dose Advisor developed for T2D [2], have shown potential in optimizing pharmacological interventions by customizing drug and dose recommendations to individual patient profiles. Similar approaches could be valuable in improving glycemic management in GDM cases, enhancing personalized care in postpartum treatment, drug discovery with therapy, and reducing long-term risks of developing chronic diseases in general.

XSL•FO
RenderX

Despite growing interest in AI-driven clinical support, current models often struggle to integrate diverse, multisource medical data into actionable insights, especially in conditions such as GDM, where missing information and diagnostic delays contribute to less desirable outcomes. These limitations are particularly prominent in resource-constrained settings, where systemic challenges, such as insufficient screening tools, lack of standardized care protocols, and limited provider training, complicate effective diagnosis and treatment [8,13]. As a result, the timely and effective treatment of GDM remains difficult, further endangering maternal and fetal health.

In such contexts, the unavailability of specialized professionals, economic constraints, and cultural challenges also influence treatment adherence and engagement [14,15]. The limited awareness between both the public and health care providers continues to contribute to improper management of GDM [16], reinforcing the urgent need for robust, context-sensitive clinical decision support [17,18].

To address these gaps, we propose a novel solution using specialized GenAI techniques for GDM management. Specifically, we develop a proof-of-concept (PoC) of a clinical support system that uses a knowledge graph (KG) supporting a local large language model (LLM). This system extracts and integrates intervention strategies from peer-reviewed research to support physicians in making contextually relevant treatment decisions.

Standalone local LLMs, however, face known limitations, including hallucinations and reduced reliability when handling domain-specific, complex queries [19]. To address these issues, we introduce a retrieval-augmented generation (RAG) mechanism that improves the accuracy and relevance of outputs by supplementing the LLM with contextual data [20,21]. This hybrid approach could elevate the clinical utility of GenAI for complex, low-resource health care scenarios such as GDM.

By generating structured, evidence-informed recommendations in real time, our system lays the foundation for scalable and explainable AI support tools customized to maternal health. The following section reviews previous ML and LLM-based approaches to GDM detection and prediction, positioning our work within this evolving research landscape. It is stated at this juncture that while the distinction between LLMs and local LLMs is clear, it is less so between local LLMs and small language models (SLMs). The prototype developed in this study assumed a local LLM architecture but could be repurposed as SLMs, particularly in resource-constrained locations of the Global South. A concise feature comparison of LLMs, local LLMs, and SLMs is provided in Textbox 1.

Textbox 1. Feature comparison of large language models, local large language models, and small language models.

Large language models

Large language models (LLMs) are typically based on deep learning, trained on massive amounts of text and increasingly multimedia data to understand, generate, and manipulate human language. LLMs work by learning to predict the next word in a sequence based on the context of the input prompt, using billions of parameters to refine these predictions. They excel at natural language processing tasks such as text completion, translation, summarization, question-answering, and content generation.

Local LLMs

Local LLMs run inside the private data center of an entity or organization. Local LLMs are fine-tuned with the organization's data (eg, patient records or standard rules) and can provide specific context to a query or prompt that general-purpose chatbots cannot or should be legally allowed to deliver. Particularly in the domains of sensitive and confidential data (such as a patient's medical conditions), such prompts may have to be subject to rigorous access, authentication, and accounting controls.

Small language model

A small language model is designed to understand and generate natural language, similar to LLMs, but on a much smaller scale, with fewer parameters and a simpler architecture. Small language models are optimized for efficiency and can be deployed on resource-constrained devices like smartphones or local servers, offering benefits such as faster training and execution, lower energy consumption, and improved privacy by allowing for on-device processing and less reliance on cloud connectivity. A use case could be first responders in emergency room situations.

Recent advances in ML have shown promise in improving the early diagnosis and personalized management of chronic conditions such as GDM. These models identify high-risk individuals during pregnancy, customize treatment plans, and ultimately enhance maternal and neonatal health outcomes. Several studies have developed ML algorithms that account for demographic variations, for example [22,23], present models customized to Asian women [10] used decision trees and ensemble learning for early GDM detection, reporting high sensitivity and specificity. However, these models often fail to capture the full complexity of GDM-related factors.

The efforts to improve model interpretability include research, such as meta-reviews of clinical studies on complications during pregnancy and their treatments [24], on clinically explainable ML approaches for blood glucose monitoring [25,26], and the use of extreme gradient boosting to identify key risk factors [27]. However, several studies [25,26,28,29] note limitations in integrating high-quality datasets, supporting real-time interventions, or embedding models within clinical systems. Table 1 presents these representative models, underscoring the trade-offs between accuracy, interpretability, and practical usability.

XSL•FO
RenderX

Table . Representative research deep learning or machine learning models for predicting gestational diabetes mellitus.

Study	Year	Model	Key contributions and limitations
Kokori et al [22] and Kumar et al [23]	2024	Demographic-specific ML[a] model	• KCs[b]: Accurate predictions for specific demographics (Asian women). • Limits: Limited integration into health care systems.
Kurt et al [10]	2023	Decision trees and ensemble	• KCs: High sensitivity and specificity. • Limits: Fails to capture all GDM[c]-related factors.
Wu et al [29]	2024	Clinically interpretable ML	• KCs: Emphasized interpretable models for GDM. • Limits: Limited real-time application.
Wu et al [25]	2022	ML-based models	• KCs: Importance of high-quality datasets. • Limits: Lacks interpretability and integration.

[a]ML: machine learning.

[b]KC: key contribution.

[c]GDM: gestational diabetes mellitus.

These limitations highlight the need for models that go beyond static risk prediction to support context-aware clinical decision-making. In this regard, LLMs offer transformative potential as they generate patient-specific recommendations by synthesizing heterogeneous clinical data. When augmented with retrieval techniques, such models become more effective.

Several recent studies have discussed the expanding role of LLMs across health care domains[30]. For example, an AI system developed for liver diseases [31] provided personalized treatment strategies that improved diagnostic outcomes. Graph-based retrieval-augmented generation (GraphRAG) integration has shown benefits in nephrology by increasing output precision and reliability [20], while LLMs have supported psychotherapy automation [32] and administrative workload reduction in personalized medicine [33]. Some of these use cases are captured in Table 2, reinforcing the applicability of RAG-augmented LLMs in clinical practice.

Table . Representative use cases of artificial intelligence in clinical health care.

Study	Year	Model	Key contributions
Ge et al [31]	2024	AI[a] model for liver diseases	Enhanced diagnostic accuracy and patient management tailored for liver diseases.
Ong et al [34]	2023	Clinical decision support system	Improved clinical decision-making with RAG[b]-enhanced LLMs[c], offering precise predictions and treatments.
Miao et al [20]	2024	LLM-RAG for nephrology	Improved accuracy and reliability in nephrology advice by integrating RAG with LLMs.
Stade et al [32]	2024	LLMs in psychotherapy	Explored the potential of LLMs to support and potentially automate aspects of psychotherapy.
Tripathi et al [33]	2024	Personalized medicine AI model	Demonstrated how LLMs can automate administrative tasks, reducing clinicians' workload from electronic medical records.

[a]AI: artificial intelligence.

[b]RAG: retrieval-augmented generation.

[c]LLM: large language model.

XSL•FO
RenderX

Noting the above, this paper proposes a novel architecture for GDM care that integrates (1) a *local LLM* for domain-specific control and privacy, (2) an *RAG* engine for contextual grounding, and (3) a *domain-specific KG* to capture interrelated medical evidence.

This combination enables real-time generation of explainable, evidence-informed treatment recommendations for GDM management, even in resource-constrained settings. As compared with previous studies, such as those by Nambiar et al [2] and Tripathi et al [33], which focused on general dosing automation or task simplification, this study addresses a critical gap: the need for adaptive, fine-grained, and explainable intervention support in the prenatal context.

From a technical standpoint, our contributions are (1) the construction of a GDM-specific KG derived from peer-reviewed literature; (2) the use of RAG-enhanced local-LLMs to retrieve, contextualize, and generate targeted care pathways; and (3) a PoC system architecture that is interpretable, domain-grounded, and designed for offline, privacy-preserving environments.

The PoC will support timely intervention and align with the practical realities of underserved clinical contexts; consider the plight of a rural doctor in the Global South, where internet connectivity, specialist clinician availability, and cutting-edge expertise may be limited. It represents a step toward deploying technically robust and clinically meaningful AI to applications of acute need.

Following this introduction, the remainder of this paper is organized as follows. The next section addresses the methods, and specifically, a description of developing design artifacts for a PoC. In the Results section, we put the system through simulated scenarios and test the responses for accuracy, bias, and performance benchmarking. In the Discussion section, we present the principal findings along with an analysis of key contributions of the research. The paper ends with a section on Conclusions, which also covers limitations and suggestions for further research.

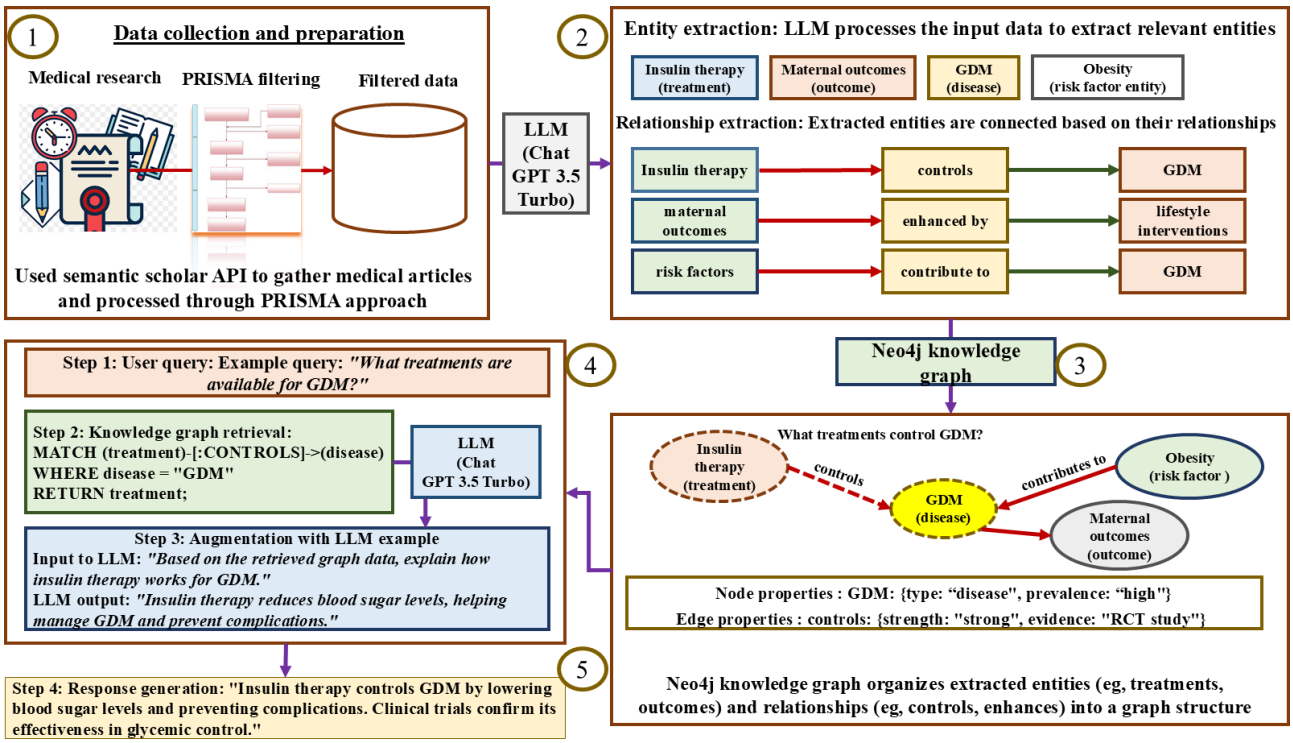
Methods

Prototyping a PoC

Health care professionals, particularly those in densely populated and resource-constrained regions of the Global South, often face significant challenges in accessing timely, evidence-based medical insights. Attending training sessions or reviewing vast volumes of literature under time pressure is impractical, especially in scenarios where specialist expertise or standardized guidelines are lacking. Our approach uses computational methods to extract, structure, and contextualize medical knowledge using GenAI and KG technologies to address this need.

Our primary objective was to develop a PoC of a clinical AI assistant that would support the management of GDM. This GraphRAG-based architecture combines entity extraction from published research, KG construction, and RAG to generate clinically grounded, context-aware responses. As illustrated in Figure 2, the PoC framework follows a 5-stage pipeline.

Figure 2. Process flow of the proposed graph-based retrieval-augmented generation approach, showing data collection, entity extraction, knowledge graph construction, and retrieval-augmented generation for AI-assisted clinical support for gestational diabetes mellitus. API: application programming interface; GDM: gestational diabetes mellitus; LLM: large language model; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



XSL•FO
RenderX

- Data collection and preparation: The Semantic Scholar API retrieved relevant research articles on GDM interventions. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)–guided filtering process was applied to ensure that inclusion criteria were met, resulting in a refined corpus of 1212 high-quality articles.
- Entity extraction: Using GPT-3.5 Turbo (OpenAI) and few-shot prompting, entities such as treatments, outcomes, risk factors, and disease indicators were extracted from full-text articles. Semantic consolidation (eg, grouping "low-carb diet" and "reduced carbohydrate intake") ensured terminological consistency.
- KG construction: Extracted entities and their relationships were encoded into a Neo4j graph database. The graph allowed efficient traversal of clinical pathways, such as connecting interventions to outcomes and risk profiles. Each node and edge pair was annotated with medical metadata, such as intervention strength, evidence level, or prevalence.
- Query processing and graph retrieval: When a user query is submitted (eg, "What treatments control GDM?"), the system was designed to retrieve relevant subgraphs using Cypher queries. These results are then passed to the LLM for augmentation and contextual response generation by incorporating patient records.
- Response generation: The final output is a clinically coherent and relevant response integrating retrieved evidence and a generative explanation. For example, based on retrieved data, the model might respond: "Insulin therapy controls GDM by lowering blood sugar levels and preventing complications." If asked why, the system might explain: "Insulin enables glucose uptake by cells throughout the body, particularly muscle and fat cells, by facilitating glucose transport across cell membranes. Without adequate insulin, glucose accumulates in the bloodstream while cells are starved of this essential energy source."

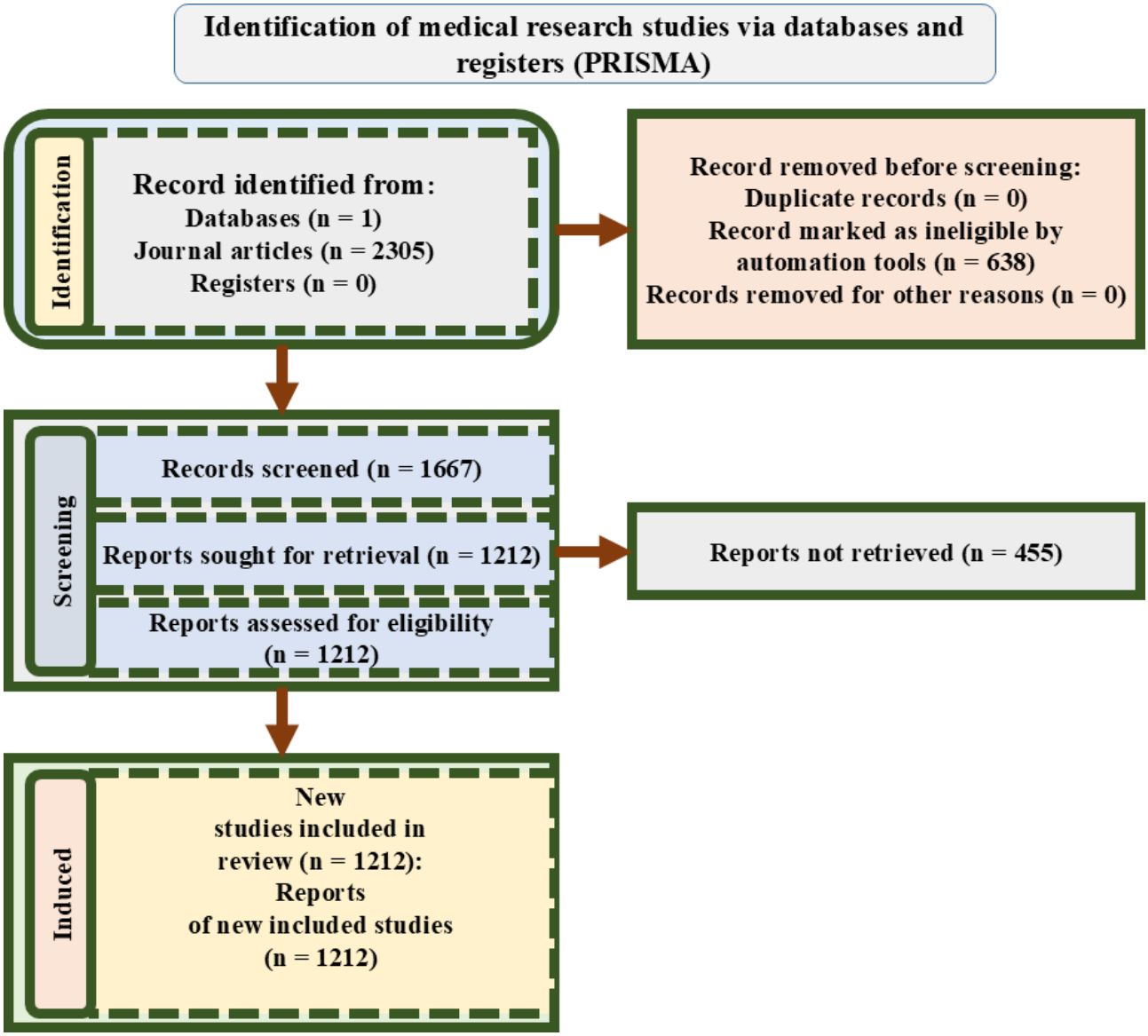
This multistep process would allow the system to access reputable and current medical research to produce explainable, evidence-grounded outputs for clinical decision support. Each component of this workflow is further detailed in the following subsections.

Data Collection

To develop a high-quality domain-specific KG for GDM, we conducted a systematic search using the Semantic Scholar API [35], a widely used biomedical research platform. The query term "gestational diabetes interventions" was selected to target studies focused on treatment strategies and clinical outcomes. The search was restricted to articles published between January 2000 and May 2024, to cover both foundational and contemporary research. The data collection and filtering process adopted PRISMA guidelines, as illustrated in Figure 3.

XSL•FO
RenderX

Figure 3. PRISMA flow diagram showing the systematic data collection and filtering process, detailing identification, screening, eligibility assessment, and inclusion of research articles for knowledge graph construction.



- Identification: The initial search produced 2305 journal articles. No records were found from registers. Automated filters removed 638 ineligible records based on metadata mismatches or irrelevant domains. No duplicate entries were detected.
- Screening: The remaining 1667 articles were screened by 2 reviewers (FR and SB) based on titles and abstracts. This stage ensured that only articles related to GDM diagnosis, treatment, management, or intervention outcomes were retained.
- Eligibility: A total of 1212 full-text articles were deemed eligible based on the inclusion criteria. Articles were excluded at this stage (n=455) due to full-text unavailability, access limitations, or insufficient clinical relevance.
- Inclusion: The final corpus consisted of 1212 peer-reviewed studies, all of which were used to extract entities and construct the GDM-focused KG.

While Semantic Scholar provided comprehensive coverage and metadata-rich access, reliance on a single source introduces potential limitations, such as limited representation of non-English or region-specific research and sensitivity to keyword variations. Future work could explore multilingual database integration and broader query strategies to reduce potential selection bias.

Nonetheless, for developing our PoC, the selected dataset offered sufficient diversity and clinical validity to enable meaningful experimentation and system development.

Entity Extraction

Following the curation of the GDM research corpus, the next step involved extracting clinically relevant concepts, including treatments, risk factors, and outcomes, from the published research. This process was executed using OpenAI's GPT-3.5 Turbo 16K API [36], which supported advanced natural language processing for domain-specific knowledge extraction. Rather than relying on pretrained biomedical ontologies, we adopted a lightweight prompting-based approach aligned with our PoC's experimental and modular goals.

XSL•FO
RenderX

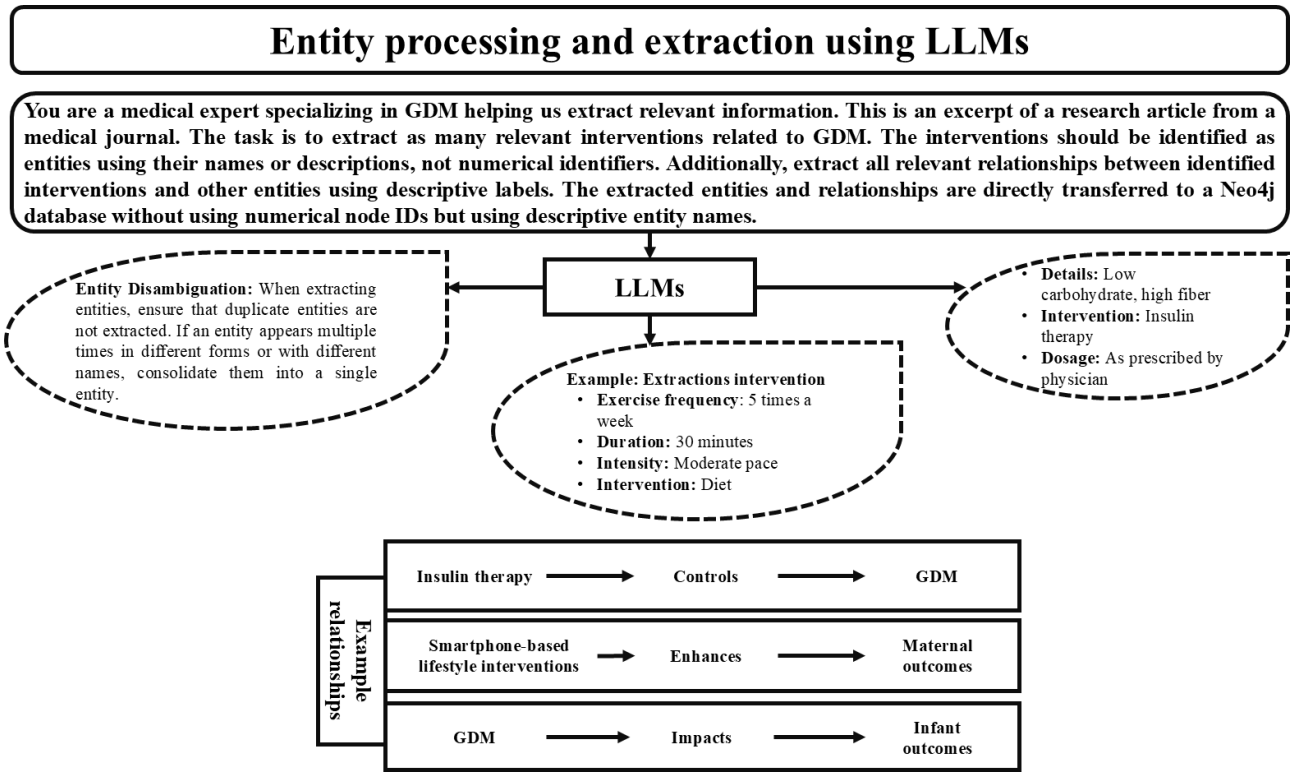
A few-shot prompting strategy was applied to guide the language model in identifying and structuring entities of interest in a usable format. Guided by 3 medical doctors, the prompts were manually engineered to show expected outputs, such as intervention types (eg, insulin therapy, diet, and physical activity), intervention parameters (eg, frequency, duration, and dosage), and associated maternal and infant outcomes. This enabled the model to consolidate synonymous or semantically related expressions (such as "low carbohydrate diet" and "reduced carb intake") into a unified entity representation. The same prompts also encouraged disambiguation of overlapping terms and discouraged the duplication of entities across articles.

The outputs were parsed into structured formats, which included both individual entities and the semantic relationships among them, for example, linking "insulin therapy" as a treatment that "controls" GDM, or connecting "smartphone-based lifestyle interventions" to enhanced "maternal outcomes." These entities and their connections were then directly integrated into the KG in the next stage of development.

This stage of entity extraction was led by the coauthor (FR), who specializes in bioinformatics and uses a technique we describe as "medical prompt engineering." The objective was to simulate how future clinical AI assistants might extract structured knowledge from unstructured medical literature autonomously. However, we acknowledge that such extractions would require validation by specialist health care professionals to ensure accuracy and reliability for clinical deployment.

The overall entity extraction workflow, including prompt design, model guidance, semantic structuring, and preparation for graph integration, is visualized in Figure 4.

Figure 4. Entity extraction workflow using large language models. The diagram is an example of the process for extracting interventions, risk factors, and relationships, which produces structured and context-aware knowledge representation for gestational diabetes mellitus management. GDM: gestational diabetes mellitus; LLM: large language model.



Construction of the KG

Upon completion of the entity and relationship extraction, the structured data were integrated into a KG using Neo4j, a widely used open-source graph database [37]. Neo4j is optimized for representing interconnected biomedical data, making it well-suited for capturing the multifactorial nature of GDM management, which involves dynamic relationships between interventions, risk factors, outcomes, and complications [38].

The KG construction process involved linking each extracted entity, such as insulin therapy, dietary strategies, or risk factors like obesity, to its semantically relevant mappings using directional edges labeled with relationship types (eg, "controls," "contributes to," and "enhances"). Each node was annotated with descriptive labels and properties derived from the literature, and relationships were encoded with metadata such as source references or study types, when available.

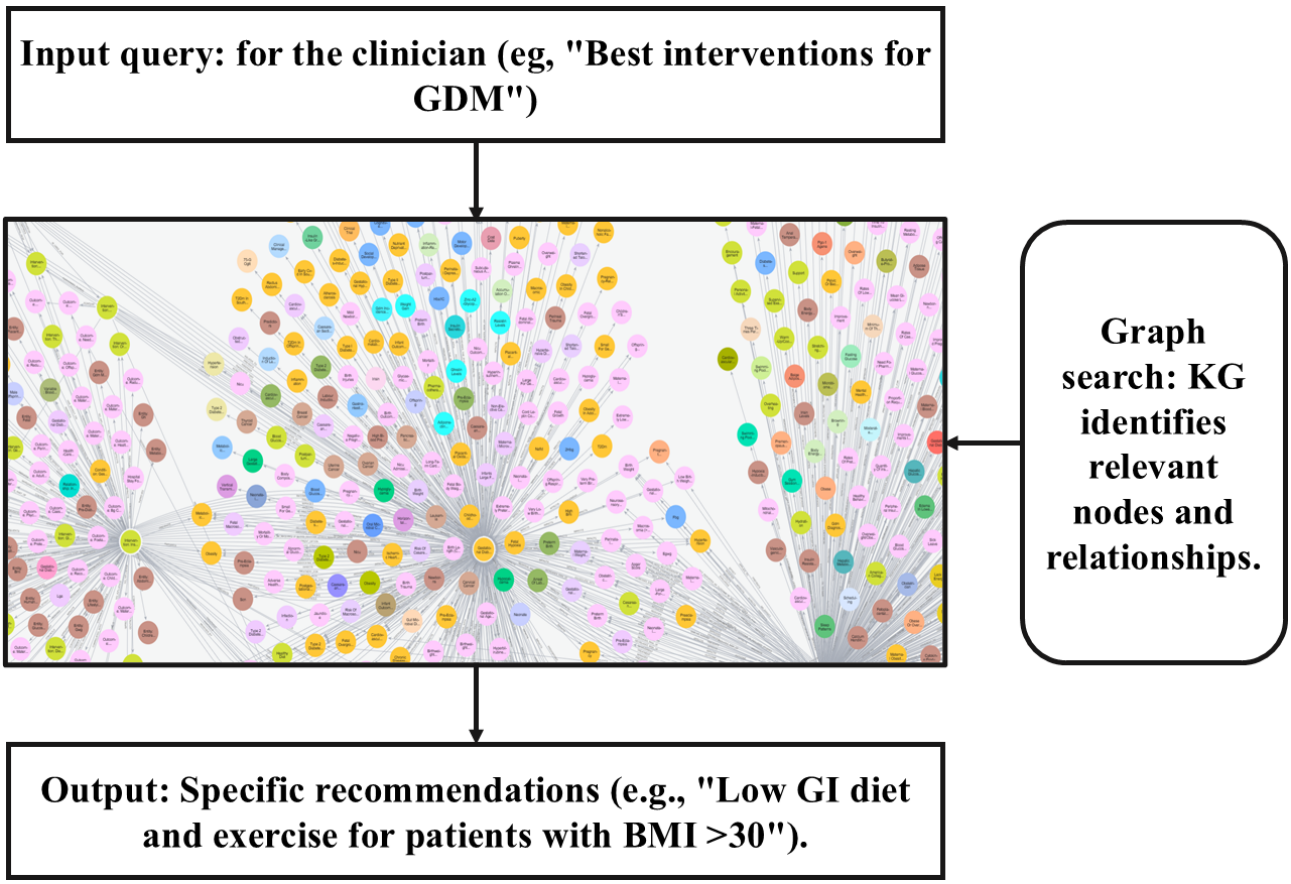
All nodes and edges were imported into Neo4j through a structured ingestion pipeline, enabling clinicians or researchers to query the KG using the Cypher query language. This functionality allowed for exploratory clinical queries, such as identifying interventions most frequently associated with improved maternal outcomes in high-risk GDM cases or tracing evidence paths for specific treatment combinations.

The resulting KG facilitated context-aware clinical decision support by surfacing specific evidence-informed insights. For example, a clinician's query, such as "What are the best interventions for GDM in patients with a BMI over 30?" could retrieve targeted graph segments linking relevant interventions

XSL•FO
RenderX

(eg, low glycemic index diet and structured exercise regimens) to outcomes validated in the literature. This dynamic capability is depicted in Figure 5, which illustrates a representative graph traversal initiated by a clinician's question, leading to personalized treatment recommendations based on the structural relationships captured in the KG.

Figure 5. Knowledge graph–powered clinical support system for gestational diabetes mellitus. The graph-based search retrieves relevant interventions and relationships, giving treatment recommendations. GDM: gestational diabetes mellitus; KG: knowledge graph.



The KG serves as the core reasoning backbone of the prototype clinical assistant, consolidating distributed medical evidence into a queryable visual knowledge substrate that can be updated as new medical evidence emerges.

KG-Based RAG

To enhance the clinical utility of the constructed KG, we then implemented an RAG approach [39]. This hybrid architecture combines traditional retrieval mechanisms with generative LLMs to produce contextually grounded and medically sound responses. In clinical settings, where decision-making depends on subtle interpretation and evidence-based insights, this integration mitigates the limitations of standalone generative systems like SLMs.

While LLMs, such as ChatGPT (OpenAI), can produce fluent and context-aware responses, they are prone to hallucinations, outdated knowledge, and domain-specific inaccuracies [19]. Conversely, RAG addresses these gaps by coupling LLMs with reputable (peer-reviewed) external knowledge sources. For example, no medical claim, such as bleach being a valid treatment for COVID-19, would have gone into the KG. In our PoC, entity-aware retrieval from the Neo4j-based GDM KG provides factual context, which the LLM then uses to generate a tailored response. This integration significantly improves factual grounding and interpretability, essential in critical domains, such as maternal health [20,21].

Using the PoC follows a 5-stage pipeline, visualized in Figures 6 and 7. Beginning with an initial clinical query, the system encodes the user input and dynamically retrieves semantically matched information from the KG. This process accounts for risk factors, interventions, and patient-specific context, including medical records and socioeconomic profiles, thereby aligning output with real-world variability in treatment planning.

XSL•FO
RenderX

Figure 6. End-to-end process flow of the graph-based retrieval-augmented generation solution. The pipeline processes medical literature and patient data, integrating them into a structured knowledge graph for AI-driven clinical decision support. GDM: gestational diabetes mellitus; KG: knowledge graph; LLM: large language model.

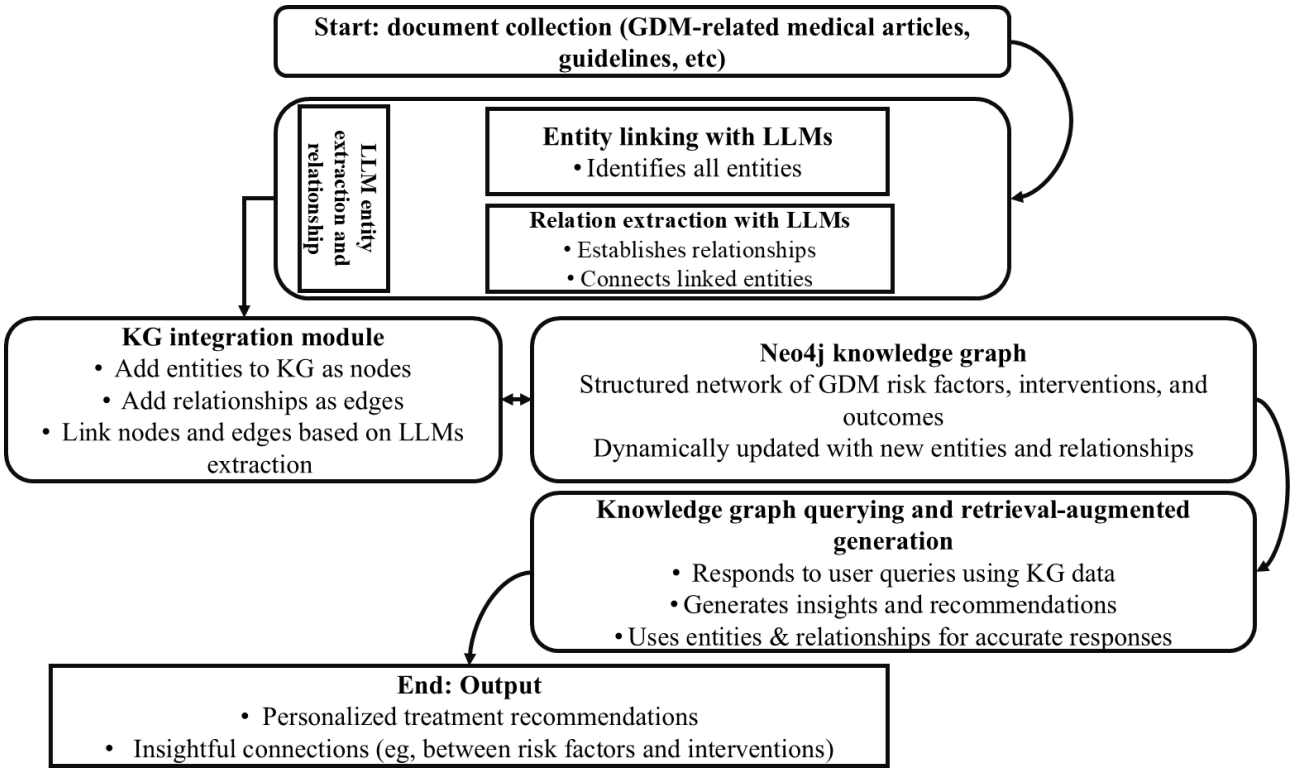
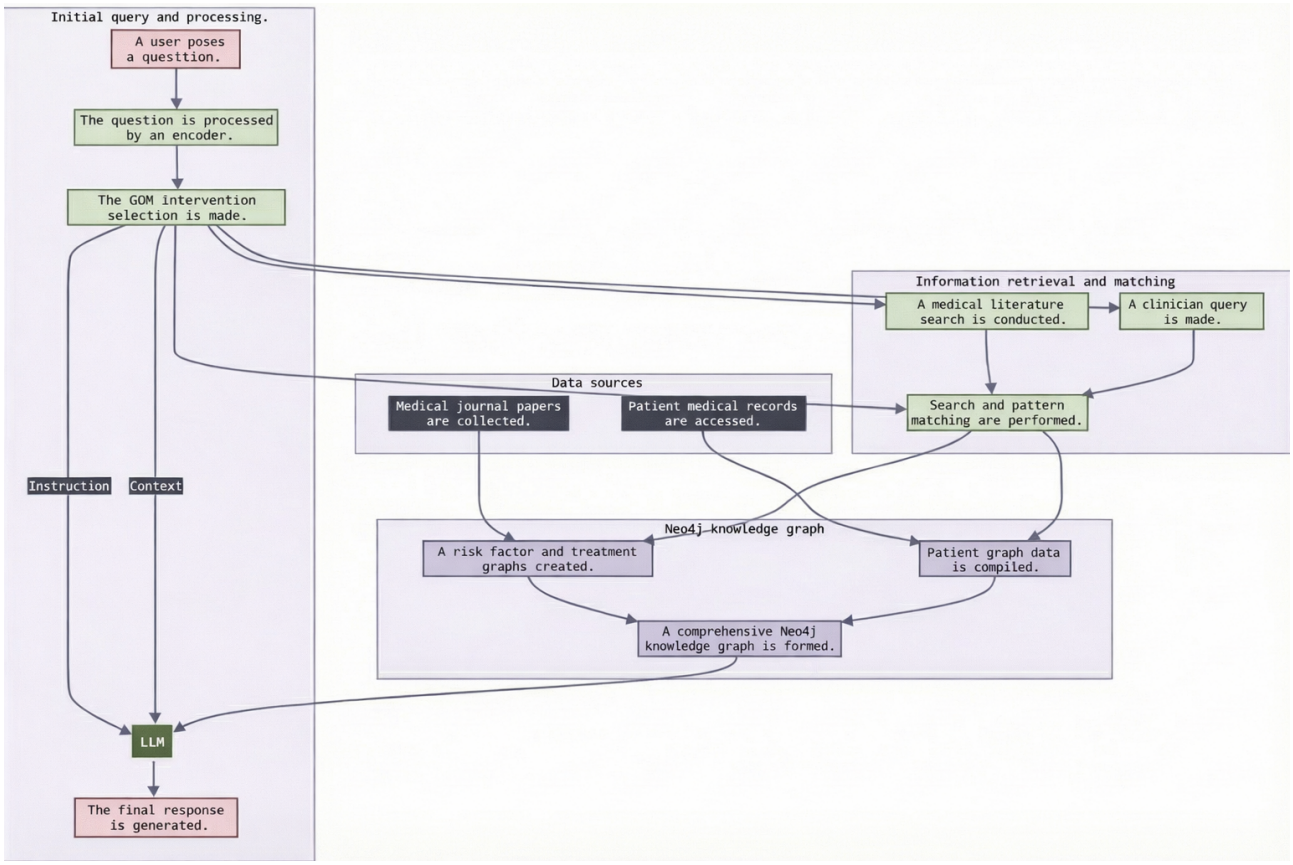


Figure 7. Structured retrieval and response generation process in graph-based retrieval-augmented generation. The diagram shows how clinician queries interact with medical knowledge sources, pattern matching, and graph-based retrieval to enhance artificial intelligence–generated responses. GDM: gestational diabetes mellitus; LLM: large language model.



XSL•FO
RenderX

Figure 6 shows the underlying LLM-KG pipeline, including entity extraction, relationship linking, and graph query generation. Figure 7 offers a complementary perspective by emphasizing end-to-end data flow, from patient query and literature matching to LLM response generation, thus highlighting how both structured (graph-based) and unstructured (textual) data are integrated to yield context-aware, personalized responses.

Although Figures 6 and 7 present a simplified overview of system functionality, the development process required iterative prompt engineering, guided tuning, and manual validation to align LLM outputs with the domain-specific vocabulary and relationships obtained from GDM research literature [32,33]. This iterative refinement helped ensure that the GraphRAG PoC consistently produces clinically meaningful recommendations rooted in the KG, avoiding spurious correlations and unverified claims.

Evaluation Framework and Metrics

The evaluation of the GraphRAG-powered local LLM for GDM was conducted through a structured framework designed to assess both technical performance and clinical relevance. Applications of AI in health care require rigorous validation beyond prompt engineering. This study used a multidimensional evaluation process using a combination of quantitative metrics and clinician-generated prompts.

Evaluation Objectives

The primary objective of the evaluation was to measure the effectiveness of the proposed PoC in three "fit for purpose" criteria: (1) generating clinically relevant**,** context-aware responses to queries on GDM management; (2) comparing its performance against widely used open-source LLMs in terms of accuracy and interpretability; and (3) assessing whether the retrieval-augmented approach of GraphRAG significantly improves response quality in medical decision support. These criteria reflect the critical nature of clinical decision-making, where AI-generated content's clarity, accuracy, and contextual relevance directly affect patient safety and clinical outcomes.

Testing Environment

The evaluation was conducted in a simulated environment, without the involvement of live patients or human participants. The GraphRAG-powered local LLM was deployed on an offline computing environment, ensuring that no external API calls or third-party cloud services influenced the test outcomes. The KG was prepopulated with medical research articles, as described in the "Prototyping a PoC" section, and served as the contextual knowledge base for all retrieval-augmented queries.

Prompt Design and Benchmark Models

The prompts used in the evaluation were carefully crafted to simulate realistic clinical and layperson queries. These prompts were generated from two user groups: (1) *laypersons* represented by 5 contributors (the authors) simulating patient queries, verified for clarity and simplicity; and (2) *clinicians* comprising 2 general practitioners (GPs) and 1 specialist physician, who created queries based on typical clinical decision-making scenarios.

Furthermore, 2 independent medical practitioners reviewed all prompts to ensure clinical relevance (were the prompts aligned with real-world GDM management scenarios?) and content clarity (did the prompts avoid ambiguous phrasing or unrealistic edge cases?)

The GraphRAG system was then benchmarked against 3 open-source LLMs commonly used in medical AI research. The comparison is intended to analyze the performance of a domain-augmented local model (our PoC) against both general-purpose and specialized health care LLMs.

- ChatGPT [36]: A versatile, general-purpose LLM.
- Claude [40]: Known for generating coherent, contextually rich responses.
- BioMistral [41]: A domain-specific medical LLM optimized for health care contexts.

Our benchmarking compares the GraphRAG-enabled local LLM against the above 3 LLM models to assess clinical relevance, contextual accuracy, and terminological consistency. These models were selected based on availability, health care domain relevance, and ease of integration into our evaluation pipeline. While we acknowledge the increasing prevalence of open-source LLMs such as LLaMA 3 (Meta AI), due to hardware compatibility constraints and inference framework differences at the time of testing, we could not integrate LLaMA 3 within the test environment. LLaMA 3 and other emerging open-source models, such as Mistral 7B (Mistral AI) and Phi-3 (Microsoft), should be included in future benchmarking updates to expand our comparative analysis, which is suggested as future work.

Evaluation Metrics and Rationale

Following established practices in evaluating health care AI models [42,43], we used 5 complementary metrics, each addressing a distinct dimension of AI-generated response quality. These are presented in Table 3.

XSL•FO
RenderX

Table . Metrics and their clinical significance in evaluating artificial intelligence–generated responses.

Metric	Purpose	Significance
Relevance score	Measures alignment between response content and user query.	Critical for clinical decision support, where irrelevant or off-topic answers compromise safety.
BLEU[a] score	Evaluates syntactic similarity and phrase structure match against reference answers.	Ensures AI[b] responses replicate validated medical language without distortion.
Jaccard similarity	Quantifies overlap in key medical terms between model response and reference.	Captures preservation of clinical terminology essential in GDM[c] management.
BERTScore	Assesses semantic similarity using deep contextual embedding.	Evaluates whether model responses capture the intended clinical meaning beyond surface text.
METEOR	Evaluate fluency and coherence in response generation.	Ensures clarity and interpretability for both clinicians and patients.

[a]BLEU: bilingual evaluation understudy.

[b]AI: artificial intelligence.

[c]GDM: gestational diabetes mellitus.

Together, these metrics comprehensively address the precision, contextual relevance, and interpretability of an AI model's outputs, which are key requirements for clinical use cases.

Evaluation Process

The evaluation adopted the following steps:

First, each LLM, including GraphRAG, was presented with the same curated set of 20 prompts (10 from simulated layperson queries and 10 from clinicians), covering core aspects of GDM management, such as risk factors, diagnostics, treatment, and complications. The 5 coauthors (EE, FR, SB, AN, and RS) jointly drafted the layperson prompts, while clinical prompts were contributed by 2 practicing GPs and reviewed by a third medical specialist.

Second, the system's responses were compared against reference answers, curated from clinical guidelines and expert consensus statements.

Third, evaluation was conducted in a zero-shot retrieval-augmented setting. No supervised training or fine-tuning was performed. The local LLM operated on a preconstructed KG as the contextual grounding source.

Fourth, automated evaluation metrics (bilingual evaluation understudy [BLEU], Jaccard Similarity, BERTScore, and METEOR) were computed using standard natural language processing evaluation libraries. These scores reflect surface-level accuracy, overlap in medical terminology, and semantic similarity.

Fifth, manual relevance scores were assigned by 2 independent medical reviewers on a 1‑5 scale, based on clinical applicability, specificity, and usefulness of responses.

Finally, results were averaged across all prompts and models and reported for comparative analysis in the Results section. While performance scores are high (eg, BLEU=0.99 approximately), this reflects a small, curated test set and should not be considered generalizable. CIs and interrater agreement were not calculated in this phase of the research.

Benchmarking Scope and Qualifications

The evaluation was designed to show the technical feasibility and domain relevance of the GraphRAG framework, rather than to establish clinical deployment readiness for deployment. Consequently, the following qualifications would apply:

First, all responses were evaluated in a simulated, offline environment without involvement of human patients, real-time electronic health record data, or live clinical workflows.

Second, no supervised training or dataset splitting was involved, as the system uses RAG rather than end-to-end training. All prompts were presented statically to each LLM.

Third, as recorded in our research logs, the KG was constructed from a curated corpus of 1212 peer-reviewed, English-language articles on GDM interventions, extracted via Semantic Scholar API (2000‑2024). The KG contains approximately 2750 nodes, 5800 edges, and 18 entity types, including risk factors, therapies, dietary interventions, and outcomes.

Fourth, the evaluation prompt set, while medically validated, remains small and nonrandomized. No demographic stratification, multilingual testing, or subgroup fairness analysis was performed.

Fifth, performance metrics assessed linguistic and contextual quality only. There has been no empirical validation of clinical efficacy, patient safety, or decision-making utility.

Finally, future iterations should expand prompt diversity, compute interrater reliability scores, and explore prompt-based fairness auditing. Prospective clinical trials and feedback-integrated deployment pipelines are also planned.

Ethical Considerations

This study involved the development and technical validation of a PoC clinical AI assistant for GDM management. The research was conducted entirely in a simulated environment without involving human participants, personal health data, or clinical interventions. Accordingly, formal ethics board approval was not required for this PoC phase of the research study.

XSL•FO
RenderX

More specifically, this was in accordance with ethical research standards for early-stage AI system development in health care. We ensured that no human participants, no personal health data, and no real-time clinical interventions resulted from this PoC phase. While fairness across subpopulations was not evaluated in this version, future efforts shall explicitly address this dimension.

Data Source Transparency

The data used in this study were drawn exclusively from public-domain, reputable academic research, collected through the Semantic Scholar API. All articles retrieved were from peer-reviewed scientific publications, ensuring no private, sensitive, or patient-level data were accessed or processed. The use of publicly available literature aligns with ethical practices in computational biomedical research, where datasets are preferably in the public domain.

Simulated Testing Environment

The PoC was evaluated using simulated prompts designed by the research team and reviewed by independent clinicians. No real patient interactions, medical records, or clinical environments were involved in the testing. This approach was explicitly chosen to focus on the feasibility of the proposed GraphRAG-powered knowledge retrieval and response generation approach.

All comparisons against open-source LLMs (ChatGPT, Claude, and BioMistral) were also conducted offline, with no data sent to external servers during evaluation, ensuring data security and compliance with our concern that we do not train such models with our research data.

Responsible AI Development

The design and development of the GraphRAG framework adhered to ethical AI principles, emphasizing:

1. Transparency: Clear explanation of methods and evaluation.
2. Safety: Avoidance of deploying untested AI systems in live clinical environments.
3. Explainability: Use of a KG for contextual reasoning and improved interpretability.
4. Bias awareness: Although no patient data were used, future iterations will integrate fairness auditing to minimize algorithmic bias.

Fairness and Demographic Representation

The development of the PoC used a small set of curated prompts authored by the research team and clinicians. Hence, no demographic, linguistic, or regional diversity was represented in the evaluation. This limitation may impact the generalizability of the system's recommendations across patient populations. Future prototyping iterations will integrate fairness-aware evaluations**,** including prompt diversity across age, gender, geography, and language, to improve equitable performance across clinical contexts.

Results

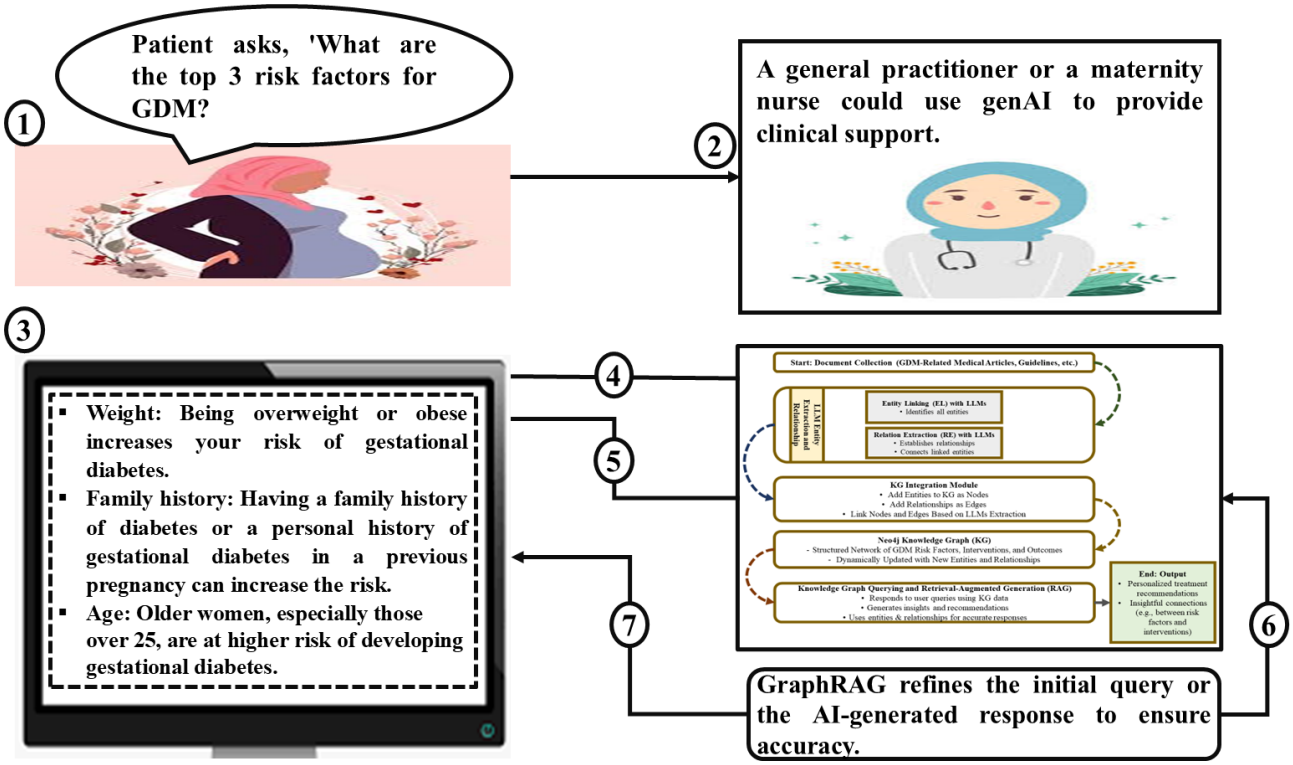
System Demonstration Scenarios

The PoC beta testing in a simulated environment highlighted the feasibility of the GraphRAG-powered clinical support system for GDM management. The PoC generated personalized, clinically relevant responses to GDM-related queries, simulating interactions between patients, health care professionals, and the system.

Figures 8 and 9 present an illustrative scenario displaying how the GraphRAG local LLM could support clinical consultations. In this example, a patient presents a question regarding the top risk factors for GDM. A health care professional, such as a GP or maternity nurse, uses the GraphRAG-enabled clinical support system to process the query into a prompt.

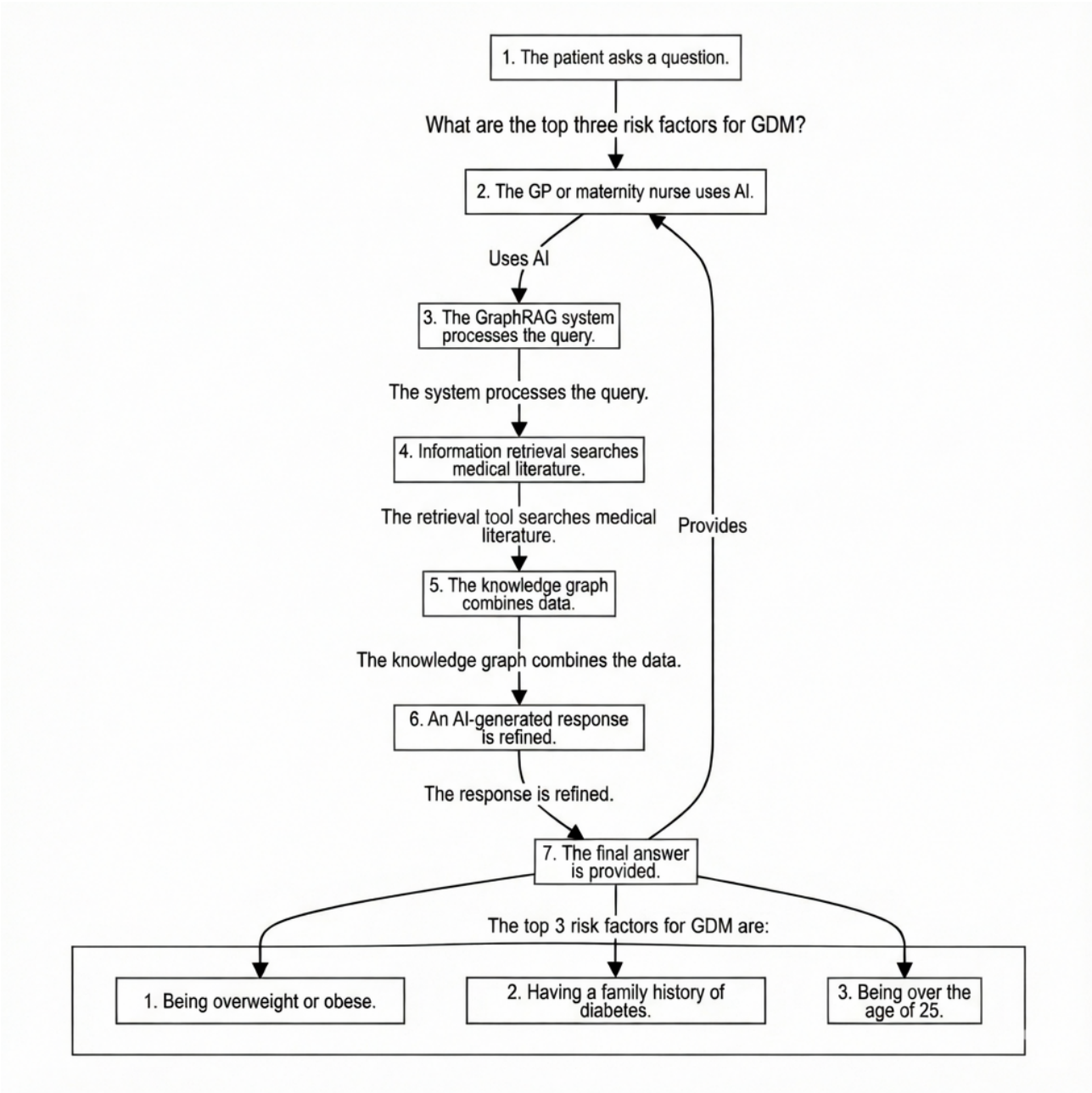
XSL•FO
RenderX

Figure 8. GraphRAG-based clinical support system for gestational diabetes mellitus - iconographic representation. AI: artificial intelligence; GDM: gestational diabetes mellitus; genAI: generative artificial intelligence.



XSL•FO
RenderX

Figure 9. GraphRAG-based clinical support system for gestational diabetes mellitus - process flow diagram. AI: artificial intelligence; GDM: gestational diabetes mellitus; GP: general practitioner.



As illustrated in Figure 8, the system processes the initial query. It generates a concise, contextually relevant response (ie, with the benefit of the patient's medical records), listing key GDM risk factors such as maternal weight, family history of diabetes, and maternal age. The process flow highlights how the system integrates domain-specific medical literature and patient-related contextual data through its underlying Neo4j KG, enabling it to deliver evidence-based, patient-centered recommendations.

Figure 9 further elaborates on the underlying process. The patient's query initiates a series of steps where the system retrieves relevant interventions and relationships from the KG. The local LLM then generates a response with context-aware medical knowledge. The final advisory to the health care professional integrates the patient's specific context and up-to-date medical research, avoiding potential inaccuracies and hallucinations.

This demonstration shows how GraphRAG can streamline clinical consultations by providing up-to-date, evidence-backed responses. Beyond basic patient interactions, the system can address more advanced clinical questions, such as identifying appropriate screening tests for GDM or evaluating intervention effectiveness. With explanation features, these capabilities support informed decision-making for health care providers and patients, enhancing the overall quality of care for GDM.

The detailed results of the system's performance in various scenarios are presented in the following 3 subsections. The first examines the system's responses to layperson prompts. The one that follows analyzes its performance with clinically oriented queries from 2 GPs. The third presents the comparative analysis of model performance using standard evaluation metrics across GraphRAG and 3 other LLMs.

XSL•FO
RenderX

GraphRAG Local LLM Response to Prospective Patients' Prompts

The GraphRAG Local-LLM generated clear, clinically relevant responses to patients' queries related to GDM management. These prompts were designed to simulate questions that a general audience might ask during health consultations with their health care provider. The system's responses highlighted its ability to simplify complex medical concepts while maintaining factual accuracy and clinical relevance.

Table 4 presents example prompts and the corresponding GraphRAG-generated responses. For example, when asked about the *top 3 factors determining gestational diabetes risk*, the system correctly identified key risk factors such as *weight, family history of diabetes, and maternal age*.

Table . Responses to patient prompts.

Prompt	Response
What are the top 3 factors determining gestational diabetes risk?	• Weight: Being overweight or obese increases your risk of gestational diabetes. • Family history: Having a family history of diabetes or a personal history of gestational diabetes in a previous pregnancy can increase the risk. • Age: Older women, especially those over 25, are at higher risk of developing gestational diabetes.
How can we safeguard a baby's health and wellness for gestational diabetes patients?	• Monitor blood sugar levels, follow a healthy meal plan, engage in regular physical activity, take prescribed medications, attend prenatal check-ups, and follow healthcare guidance.
What nutrition is recommended to prevent gestational diabetes?	• Eat a balanced diet high in fiber and low in added sugars and unhealthy fats. Include fruits, vegetables, whole grains, lean proteins, and healthy fats.
Are gestational diabetes patients more likely to have a cesarean or a standard delivery?	• Women with gestational diabetes are more likely to have a cesarean delivery due to risks like macrosomia, shoulder dystocia, and birth trauma.
What neonatal screening should a couple with a family history of diabetes undergo before planning to have a baby? Are there precautions they could take?	• Consider genetic counseling and testing. Preconception counseling can help identify risks and precautions, with personalized plans developed alongside health care professionals.

These results show the system's capacity to convey medical information in a manner accessible to nonspecialist users. The responses accurately addressed the queries using medically appropriate language, ensuring clarity for a general audience.

Responses to GP Prompts

The GraphRAG system was evaluated using clinically relevant prompts provided by GPs. These prompts simulated real-world clinical queries related to the diagnosis, management, and long-term planning of GDM. The system generated medically valid responses as assessed by the participating GPs and diabetologists, addressing both basic and advanced aspects of GDM care.

Responses to Prompts From GP 1

Table 5 presents the responses generated by GraphRAG for the first set of prompts from GP 1. These prompts covered foundational aspects of GDM diagnosis and management. For example, when asked how to diagnose GDM, the system recommended the glucose challenge and tolerance tests, reflecting standard clinical guidelines. The responses addressed both diagnostic procedures and recommended clinical management steps.

The responses provided clear, evidence-aligned answers covering diagnostic tests, management strategies, fetal monitoring approaches, and long-term complications.

XSL•FO
RenderX

Table . Responses to prompts from general practitioner 1.

Prompt	Response
How do I diagnose gestational diabetes?	Gestational diabetes is typically diagnosed through a glucose challenge or glucose tolerance test. A sugary solution is consumed, and blood sugar levels are tested. If results are high, a more detailed glucose tolerance test is done. A health care provider must confirm the diagnosis and guide management.
What measures would you advise me to take once I am diagnosed with GDM[a]?	Monitor blood sugar regularly, follow a health care–guided meal plan, engage in physical activity, attend prenatal care visits, take prescribed medications, and stay in contact with your health care team.
What fetal changes should I monitor (and how) in gestational diabetes?	Monitor fetal growth via regular ultrasounds, track amniotic fluid and fetal movements (kick counts), and consult health care providers for ongoing prenatal assessments.
What are the cutting-edge medical and non-medical interventions I could take to control GDM in pregnant mothers?	Use technologies like continuous glucose monitors and personalized insulin therapy. Also, nonmedical options such as tailored diets, physical activity, and stress reduction should be applied, ideally with specialist consultation.
What are the long-term complications for GDM mothers and their offspring?	GDM increases the risk of type 2 diabetes in mothers and raises the chance of obesity, metabolic syndrome, and diabetes in children. Long-term follow-up is essential for both.

[a]GDM: gestational diabetes mellitus.

Prompts From GP 2 and GraphRAG Local LLM Responses

Further evaluation was conducted using prompts from GP 2, which explored broader clinical topics, including screening tests, genetic factors, and emerging therapies. Table 6 summarizes the system's responses.

According to the medical practitioners, the GraphRAG system generated contextually appropriate responses, addressing established clinical practices and emerging medical topics in GDM care.

Table . Responses to prompts from general practitioner 2.

Prompt	Response
What basic screening tests for GDM[a] should I perform and monitor regularly?	Standard tests for screening and monitoring GDM include: • Glucose challenge test (GCT) – a one-hour blood test after consuming a sugary drink. • Oral glucose tolerance test (OGTT) – a follow-up fasting test with multiple blood draws. • Glycated hemoglobin (A_{1C}) – reflects average glucose levels over the past months. • Regular blood glucose monitoring – daily checks using a glucometer. Consult a healthcare provider for tailored testing schedules.
On balance, does medical evidence suggest a greater genetic influence of the father or the mother for DM[b]?	Both parents can genetically influence the risk of DM. The extent varies by DM type and individual genetic profiles. Environmental factors also play a key role.
If the patient and her husband decide to go ahead with starting a family, what medical interventions can I prescribe to alleviate the risk of GDM?	Medical interventions include: • Healthy diet and lifestyle habits. • Regular blood glucose monitoring. • Medications such as metformin, if required. • Referral to a nutritionist or diabetes educator. • The provider should personalize recommendations.
Can proper oral and dental care alleviate GDM during pregnancy? What are some effective measures?	There is no direct link between oral care and GDM prevention. However, good dental hygiene supports overall pregnancy health. Recommendations include brushing twice daily, flossing, and routine dental visits.
Are there promising stem-cell or epigenetic treatments that could ease my patient's hesitance to start a family?	Stem-cell and epigenetic research show potential, but are still in the early stages. Patients should consult reproductive specialists for the latest updates and personalized advice.

[a]GDM: gestational diabetes mellitus.

[b]DM: diabetes mellitus.

XSL•FO
RenderX

Comparative Model Performance

Overview of Benchmarking Procedures

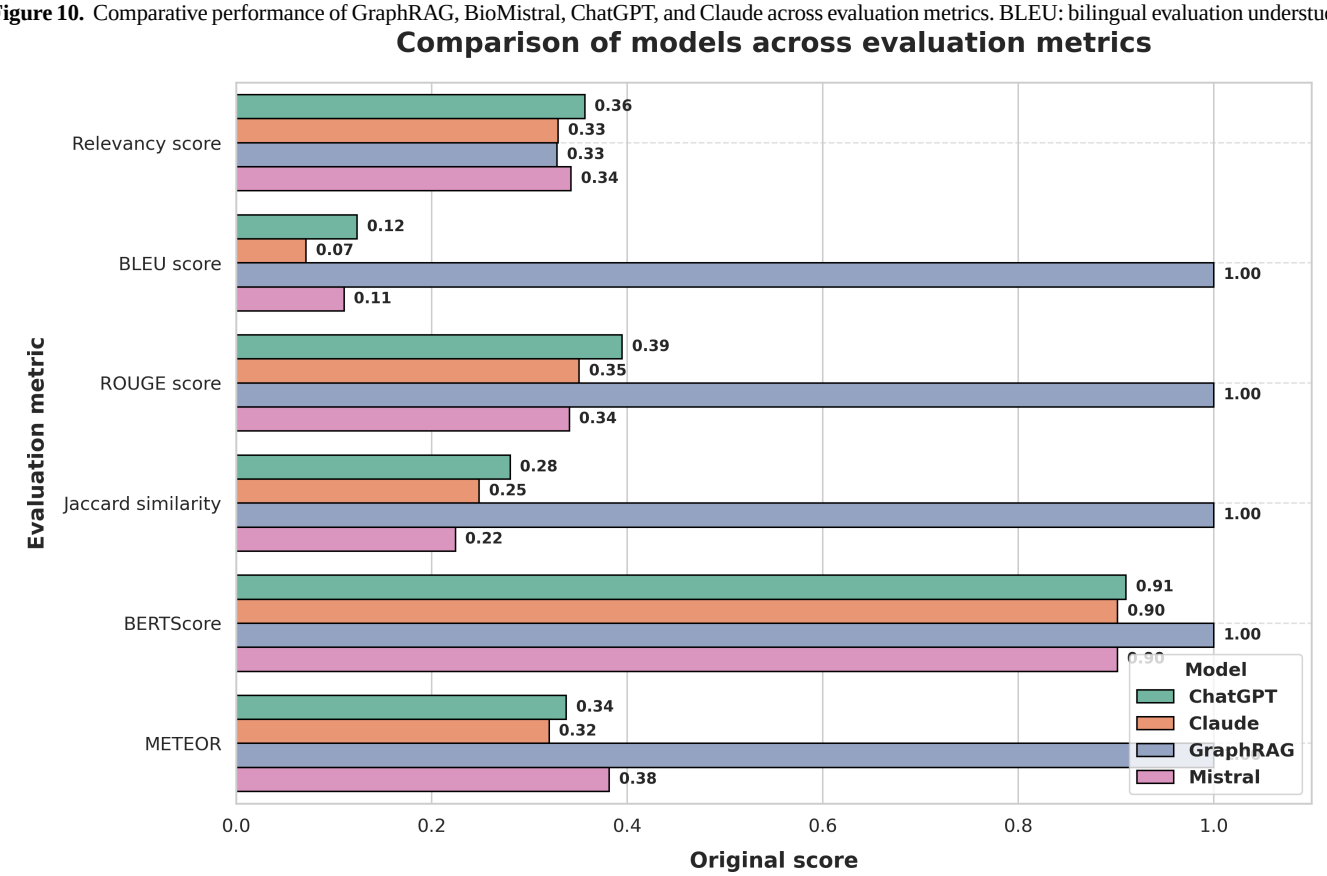
The GraphRAG system was benchmarked against 3 widely used LLMs, BioMistral, ChatGPT, and Claude, using a standardized set of clinical prompts focused on GDM management. The models' responses were evaluated using 5 quantitative metrics that assessed relevance, linguistic precision, terminology consistency, contextual understanding, and coherence.

Benchmarking Results

Figure 10 presents a comparative analysis of the models' average performance across 5 evaluation metrics. GraphRAG achieved the highest scores in BLEU, Jaccard Similarity, and BERTScore, indicating strong alignment with clinical phrasing, preservation of key medical terms, and deep contextual accuracy. Relevance Score and METEOR also reflect competitive performance across all models.

Figure 11 shows a radar chart (also known as a Kaviat diagram) of the same results, highlighting GraphRAG's balanced strengths across multiple evaluation dimensions.

Figure 10. Comparative performance of GraphRAG, BioMistral, ChatGPT, and Claude across evaluation metrics. BLEU: bilingual evaluation understudy.



XSL•FO
RenderX

Figure 11. Radar chart visualizing model performance across key metrics. BLEU: bilingual evaluation understudy.

Key Observations

1. Relevance Score: GraphRAG and BioMistral showed comparable results, aligning well with the clinical intent of queries.
2. BLEU Score: GraphRAG outperformed all other models, reflecting precise replication of validated clinical expressions.
3. Jaccard Similarity: GraphRAG highlighted superior consistency in medical terminology usage across responses.
4. BERTScore: The model achieved the highest semantic similarity, indicating deep contextual understanding.
5. METEOR: GraphRAG generated coherent and fluent responses suitable for clinical communication, comparable with ChatGPT and Claude.

These findings demonstrate the technical feasibility of the proposed GraphRAG-enabled local LLM. However, we stress that as a PoC evaluated in a simulated environment, the prototype is not ready to be deployed in real-world clinical settings. Even so, these results show that the GraphRAG approach effectively balances linguistic precision, contextual depth, and clinical relevance in GDM decision support scenarios. Besides BioMistral, ChatGPT, and Claude, new open-source LLMs such as LLaMA 3, Mistral 7B, and Phi-3 are becoming prevalent in health care AI. Although hardware and framework limitations prevented their inclusion in this study, we recognize their importance as baselines. Future work will add these models to expand our comparative analysis.

Discussion

Principal Findings

This study demonstrates that the GraphRAG-enabled local LLM consistently produces clinically relevant, contextually grounded, and medically precise responses for managing GDM. Through a rigorous benchmarking process against established open-source models, BioMistral, ChatGPT, and Claude, the GraphRAG approach achieved top-tier performance across all key evaluation metrics: BLEU Score, Jaccard Similarity, and BERTScore. The radar chart (Figure 11) illustrates well-balanced strengths of the model across multiple evaluation dimensions. Notably, first, GraphRAG matched BioMistral in

XSL•FO
RenderX

Relevance Score, indicating strong alignment with the clinical intent behind queries of the health care providers. This score reflects how accurately the system understands and responds to the specific clinical context of the user, which is critical in decision support. Second, GraphRAG outperformed all models in BLEU Score and Jaccard Similarity, showcasing its ability to reproduce clinical phrasing with syntactic accuracy and maintain consistency in key medical terminologies, a vital factor for preserving the meaning of technical medical advice. Third, it achieved a superior BERTScore, showing deep semantic understanding. This reflects the capacity of the model to generate responses that not only match expected language structures but also accurately convey complex clinical relationships within GDM care. Finally, the superior performance across these diverse metrics stems from the architectural design of GraphRAG. By integrating domain-specific KGs with RAG, the system grounds its responses in verified clinical evidence rather than relying solely on probabilistic language patterns. This integration mitigates common challenges of general-purpose LLMs, such as hallucinations and domain irrelevance, ensuring that responses are both medically accurate and context-aware.

These findings validate the core hypothesis that combining structured medical knowledge and patient-specific information with GenAI significantly enhances the quality, safety, and reliability of AI-generated clinical recommendations.

Implications for AI-Driven Clinical Decision Support

The empirical results of this study highlight the transformative potential of GraphRAG-powered local LLMs in advancing AI-enabled clinical decision support systems. This is particularly the case within specialized domains, such as GDM management, where expertise is scarce and uneven. By combining structured retrieval from a domain-specific KG with contextual generation, the system addresses the longstanding gap between unstructured clinical queries and structured evidence-based medical knowledge.

The GraphRAG approach addresses a critical limitation of general-purpose LLMs, such as ChatGPT and Claude, which often prioritize linguistic fluency over clinical accuracy. While these models can generate coherent responses, they frequently lack the domain specificity needed for accurate clinical guidance. In comparison, the responses of GraphRAG consistently align with established clinical guidelines, reflecting a deep understanding of current medical standards and practices. For example, when prompted to hear about GDM diagnosis, GraphRAG accurately recommended the glucose challenge and tolerance tests, mirroring clinical best practices. This indicates that the system is not merely generating plausible text but retrieving and contextualizing domain-specific evidence to support clinical decision-making.

The clinical utility of such contextually enriched responses is profound. In healthcare, where treatment decisions directly affect patient safety and outcomes, factual accuracy and contextual relevance are not optional but essential. The ability of GraphRAG to consistently deliver these qualities positions it as a valuable tool for supporting health care providers, particularly in low-resource or high-pressure clinical environments where access to specialist knowledge may be limited.

The practical utility of GraphRAG is further illustrated in Textbox 2, which presents a representative response to a clinically relevant query about managing insulin resistance in patients with GDM. Unlike generic language models, GraphRAG provides structured, evidence-aligned recommendations grounded in recent clinical guidelines.

Textbox 2. GraphRAG response to a clinical query on managing GDM.

Query: What are the latest recommendations for managing insulin resistance in GDM patients?

GraphRAG response: Recent studies suggest a multi-faceted approach to managing insulin resistance in GDM patients, including:

1. Personalized low glycemic index diets, emphasizing complex carbohydrates and fiber-rich foods.
2. Structured exercise programs combining aerobic and resistance training for at least 150 minutes per week.
3. Early initiation of metformin where lifestyle modifications are insufficient, as per the 2023 guidelines.
4. Continuous glucose monitoring to optimize insulin dosing and adjust lifestyle recommendations.
5. Regular screening for comorbidities such as hypertension and dyslipidemia, which contribute to insulin resistance.

The above sections highlight the ability of GraphRAG to transform unstructured clinical questions into actionable, guideline-compliant insights. By synthesizing evidence from domain-specific KGs, the system avoids unsupported claims and produces responses aligned with best clinical practices, supporting its role as a trustworthy clinical decision support tool.

Contributions to AI in Health Care

This study advances the field of health care AI by presenting a scalable, contextually enriched clinical support system specifically designed for GDM management. We believe that our key contribution lies in the system's ability to empower GPs and nonspecialist clinicians, particularly in underserved and resource-limited health care environments with limited access to endocrinology specialists and up-to-date clinical knowledge. By using a KG-driven retrieval process, the system surfaces context-specific clinical insights without requiring clinicians to conduct exhaustive manual literature reviews or consult multiple sources. Here, a word of caution is in order. We reiterate that the PoC works best as a clinical assistant; that is, a health practitioner must be in the loop. This is important given the dangers of unsupervised AI agents, which may usurp the role of a human caregiver without human oversight [44]. It is concerning that a recent, peer-reviewed (and in our view,

XSL•FO
RenderX

misguided) study actually normalizes a doctor versus machine "Turing-test of authenticity" [45].

Furthermore, this study shows domain-specific superiority over general-purpose LLMs. While models such as ChatGPT and Claude can produce coherent responses, they lack the fine-tuned contextual sensitivity and clinical precision essential for specialized health care domains. In comparison, the architecture of GraphRAG is optimized to capture the complex relationships inherent in GDM management, such as patient history with risk factors, availability of interventions, and outcome pathways for follow-up medical care, thereby enhancing both response accuracy and clinical applicability.

This study contributes to a novel retrieval-augmented GenAI architecture that translates domain-specific medical knowledge into clinically actionable insights. It serves a need; namely, access to the latest, credible medical research in time- and resource-constrained environments. In health care, timely and science-based interventions are crucial.

Technical Innovations Driving Performance Gains

The robust performance of the GraphRAG-enabled local LLM stems from the integration of 3 core technical innovations that address longstanding limitations in clinical AI systems.

First, the KG integration allows for the structured representation of complex clinical relationships between risk factors, interventions, symptoms, and outcomes. Unlike flat text embedding, the KG enables the system to reason over interconnected entities and contextual dependencies, ensuring that recommendations are grounded in the complete clinical scenario rather than isolated data points.

Second, the RAG framework of the system addresses the gap between static model knowledge and dynamic, evolving medical evidence. The system mitigates temporal gaps by integrating retrieval from an up-to-date domain-specific KG. It reduces the risk of hallucinated or outdated responses, a common flaw in general-purpose LLMs trained on static corpora.

Third, the domain-specific adaptation of the model through targeted prompting strategies and fine-tuning on GDM-related interventions enhances its ability to understand and accurately apply specialized clinical terminology in localized contexts. This adaptation ensures that the system's responses reflect the nuanced requirements of GDM management, capturing both the syntactic precision and semantic depth necessary for high-stakes clinical situations like emergency room triage.

We believe that these innovations enable the system to move beyond generic language generation, delivering interpretable, actionable, and clinically validated responses. This advancement represents a meaningful step toward reliable AI-assisted clinical decision-making, especially for chronic disease management scenarios where timely and context-aware recommendations are essential.

Conclusions

Limitations and Challenges for Clinical Deployment

While the initial results from this PoC study are promising, several critical limitations must be addressed before GraphRAG can be translated into clinical practice. Intended as a PoC, the system has not undergone field validation. Future studies involving real-world patient interactions, clinician feedback, and longitudinal follow-up are essential to establish the model's safety, reliability, and usability in live health care environments.

A second major consideration concerns data privacy and protection. Although this PoC did not involve patient-level data, real-world deployments would necessitate strict adherence to data protection frameworks. The integration of privacy-preserving learning paradigms, such as federated learning, would allow models to be trained on decentralized clinical data without exposing sensitive patient information. Complementary techniques, such as blockchain for differential privacy and secure multiparty computation, could further protect patient confidentiality.

The interpretability of AI-generated clinical responses remains a pressing challenge. While GraphRAG uses structured retrieval to enhance contextual grounding, clinicians must be able to trust and explain its outputs. Future iterations of the system should integrate explainability frameworks such as Shapley Additive Explanations or Local Interpretable Model-agnostic Explanations, enabling clinicians to trace and retrieve evidence on how specific KG pathways contribute to a given clinical recommendation.

In addition, seamless workflow integration will be critical for adoption. Clinical decision support systems must embed naturally within existing electronic health record platforms, minimizing disruption to physician workflows. Without such integration, even the most accurate systems risk being underused in clinical practice.

As with many multistage AI pipelines, GraphRAG is also subject to the risk of error propagation, where inaccuracies in earlier stages, such as entity extraction or graph construction, may be compounded in downstream response generation. While our current prompt engineering and domain-specific graph design reduce this risk, future versions will integrate intermediate validation checkpoints, feedback loops, and retrieval-failure auditing to ensure response fidelity and system transparency.

Another key limitation is the reliance on English-language peer-reviewed articles from a single aggregator (Semantic Scholar). This has excluded regional or non-English medical literature with culturally adapted GDM interventions. Future work should incorporate multilingual and regionally diverse corpora to improve the model's generalizability and contextual sensitivity, particularly in Global South health care settings.

Finally, the computational demands of GraphRAG's RAG architecture present scalability challenges. The latency and resource consumption must be optimized to support real-time inference in time-sensitive clinical settings, especially in environments where computational capacity may be limited. Addressing these challenges is essential for transitioning GraphRAG from an academic PoC to a clinically viable, ethically responsible AI system.

XSL•FO
RenderX

Broader Implications and Future Research Directions

Building on the demonstrated feasibility of our PoC, our future research agenda is designed to advance the GraphRAG framework along 2 primary axes: strategic domain expansion and core technical refinement. First, we propose to strategically adapt the framework for other data-intensive clinical areas, including cardiovascular disease, oncology, and mental health, where evidence-grounded decision support is crucial. Second, we will enhance the core retrieval engine by integrating advanced algorithms, such as contextual BM25 and embedding-based summarization, to improve precision. To improve robustness and transparency, we propose implementing new retrieval-specific metrics, such as recall and failure rates. We have established a roadmap and aim to pursue these enhancements in our next research cycle, solidifying the GraphRAG pipeline as a viable tool for real-world clinical decision support.

The legal, ethical, and intellectual property considerations will also shape future deployments. To ensure transparency and reduce legal risks, future iterations will prioritize training on open-access datasets such as PubMed Central, adhering to responsible AI development practices and open science principles.

To protect patient privacy and mitigate algorithmic bias will remain core ethical imperatives. The federated learning and anonymized blockchain solutions could support decentralized training across institutions without compromising patient confidentiality. Bias audits, fairness-aware modeling, and hallucination mitigation strategies, such as reranking retrieved evidence and diversifying training datasets, will improve the reliability and equity of the system's clinical recommendations. In such a trusted platform, integrating GraphRAG with real-time patient data could enable personalized clinical decision support, customizing recommendations to individual genetic profiles, lifestyle factors, and environmental exposures. This evolution toward precision medicine would represent a significant leap forward in AI-driven health care delivery.

To overcome the limitation of computational costs, the enhanced system will require architectural optimizations to enable scalability in resource-constrained clinical settings. Techniques such as prompt caching, adaptive chunking of graph queries, and hybrid retrieval strategies will reduce computational costs and response latency. This will support deployments in low-bandwidth environments, such as rural clinics and community health centers.

In the long term, retrieval-augmented LLMs, such as GraphRAG, are envisioned not as autonomous clinical agents but as clinical copilots, supporting, rather than replacing human clinicians. Their evaluation in live clinical workflows will be critical to determining their optimal role as decision-support systems. A reflective perspective on this motivation is presented in Textbox 3, showing the personal origins of our research question.

Textbox 3. Closing vignette on gestational diabetes.

"I do not wish to alarm you, Mrs. Sharma, but you have been diagnosed with gestational diabetes and your baby is 10 pounds at birth. Both of you need to be careful."

[Ward Nurse in Singapore's Kandang Kerbau Maternity Hospital to the mother of the last author, circa 1961]

In 2022, the mother passed away peacefully at the age of 88, her diabetes controlled with insulin injections for decades. The "baby" (the last author and principal investigator of this study [RS]) was diagnosed with type 2 diabetes at the age of 60, giving rise to our research question of whether a graph-based retrieval-augmented generation solution could change the outcome for both with timely, relevant best practices.

In closing, this paper sought to establish the feasibility of a GraphRAG-enabled local LLM architecture for generating clinically relevant, context-aware responses in the management of diseases, such as GDM [46]. By integrating domain-specific KGs with RAG, the system outperformed general-purpose LLMs across multiple evaluation metrics, offering evidence-grounded and terminologically precise clinical recommendations. While this work serves as a technical PoC, future research will need to focus on (1) prospective clinical validation involving real-time patient interaction, (2) multimodal agents to improve accessibility and cultural sensitivity, and (3) integration of explainable AI modules, such as Shapley Additive Explanations–based KG traceability, resulting in enhanced trust and transparency for the 2 key humans in the loop – the patient and her doctor. Ultimately, we believe the transformative potential of AI-powered decision support tools will personalize care and improve clinical outcomes, particularly in underserved societies.

Acknowledgments

The authors are grateful to the three medical practitioners, Dr Z (a diabetologist), Dr S (an emergency room physician), and Dr R (a general physician), for their thoughtful contribution to the prompts and for examining the GraphRAG responses for face validation. Many thanks are also due to the anonymous reviewers for guiding the revisions of the manuscript. No human subjects nor live patient data were used to develop the PoC.

Funding

This research was supported by the Research Incentive Fund Grant R22086 from Zayed University to AN, RS, and EE.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available in the GITHUB repository [46].

Authors' Contributions

AN and RS conceived the research idea and were Principal Investigators. FR and SB conducted the empirical data collection and validation that produced the GraphRAG model. EE supervised the research and subsequent reporting as co-PI. All authors contributed equally to the research, analysis, and writing of this article.

No potential conflicts of interest are reported by the authors.

Conflicts of Interest

None declared.

References

1. Sharma R, Zhang C, Wingreen SC, Kshetri N, Zahid A. Design of blockchain-based precision health-care using soft systems methodology. IMDS 2019 Dec 31;120(3):608-632. [doi: 10.1108/IMDS-07-2019-0401]
2. Nambiar M, Bee YM, Chan YE, et al. A drug mix and dose decision algorithm for individualized type 2 diabetes management. NPJ Digit Med 2024 Sep 17;7(1):254. [doi: 10.1038/s41746-024-01230-5] [Medline: 39289474]
3. Kjos SL, Buchanan TA. Gestational diabetes mellitus. N Engl J Med 1999 Dec 2;341(23):1749-1756. [doi: 10.1056/NEJM199912023412307] [Medline: 10580075]
4. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010 Mar;33(3):676-682. [doi: 10.2337/dc09-1848] [Medline: 20190296]
5. Akinyemi OA, Omokhodion OV, Fasokun ME, Makanjuola D, Ade-Ojo IP, Adeniyi AA. Screening for gestational diabetes mellitus: is there a need for early screening for all women in developing countries? Cureus 2023;15(2):e35533. [doi: 10.7759/cureus.35533]
6. Ismail L, Materwala H. IDMPF: intelligent diabetes mellitus prediction framework using machine learning. ACI 2025 Jan 30;21(1/2):78-89. [doi: 10.1108/ACI-10-2020-0094]
7. Ejaz Z, Azhar Khan A, Sebghat Ullah S, Aamir Hayat M, Maqbool MA, Amin Baig A. The effects of gestational diabetes on fetus: a surveillance study. Cureus 2023 Feb;15(2):e35103. [doi: 10.7759/cureus.35103] [Medline: 36938248]
8. Macaulay S, Dunger DB, Norris SA. Gestational diabetes mellitus in Africa: a systematic review. PLoS ONE 2014;9(6):e97871. [doi: 10.1371/journal.pone.0097871] [Medline: 24892280]
9. Stewart-Field ZA. Gestational diabetes: screening, diagnosis, treatment and management. Obstetrics, Gynaecology & Reproductive Medicine 2023 Jul;33(7):185-189. [doi: 10.1016/j.ogrm.2023.04.001]
10. Kurt B, Gürlek B, Keskin S, et al. Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques. Med Biol Eng Comput 2023 Jul;61(7):1649-1660. [doi: 10.1007/s11517-023-02800-7] [Medline: 36848010]
11. Jani S, Parikh P, Patel K, Shah A, Patel B, Rangrej R. Fetomaternal outcome in patients with gestational diabetes mellitus. Natl J Physiol Pharm Pharmacol 2023;13(3):1. [doi: 10.5455/njppp.2023.13.01032202306022023]
12. Alum E, Ugwu OPC, Obeagu EI, Uti DE, Egba SI, Alum BN. Managing the dual burden: addressing mental health in diabetes care. Elite Journal of Medical Sciences 2024;2(6):1-9 [FREE Full text]
13. Hinneh T, Jahn A, Agbozo F. Barriers to screening, diagnosis and management of hyperglycaemia in pregnancy in Africa: a systematic review. Int Health 2022 May 2;14(3):211-221. [doi: 10.1093/inthealth/ihab054] [Medline: 34432050]
14. Karakoc Kumsar A, Taskin Yilmaz F, Demirel G. Preferences of participating in diabetes screening programs for postpartum women with gestational diabetes mellitus in a university hospital in Turkey. JHR 2022 Apr 27;36(3):524-532. [doi: 10.1108/JHR-08-2020-0372]
15. Sahu B, Babu GR, Gurav KS, et al. Health care professionals' perspectives on screening and management of gestational diabetes mellitus in public hospitals of South India - a qualitative study. BMC Health Serv Res 2021 Feb 12;21(1):133. [doi: 10.1186/s12913-021-06077-0] [Medline: 33579259]
16. Nikolaidis GF, Duarte A, Griffin S, Lomas J. "Beyond the mean" in biomarkers modelling for economic evaluations: a case study in gestational diabetes mellitus. In: Baltagi BH, Moscone F, editors. Recent Developments in Health Econometrics: A Volume in Honour of Andrew Jones: Emerald Publishing Limited; 2024, Vol. 297:1108. [doi: 10.1108/S0573-85552024297]
17. Morampudi S, Balasubramanian G, Gowda A, Zomorodi B, Patil AS. The challenges and recommendations for gestational diabetes mellitus care in India: a review. Front Endocrinol (Lausanne) 2017;8:56. [doi: 10.3389/fendo.2017.00056] [Medline: 28392778]
18. Mpondo BCT, Ernest A, Dee HE. Gestational diabetes mellitus: challenges in diagnosis and management. J Diabetes Metab Disord 2015;14(1):42. [doi: 10.1186/s40200-015-0169-7] [Medline: 25977899]

XSL•FO
RenderX

19. Pal A, Umapathi LK, Sankarasubbu M. Med-HALT: Medical Domain Hallucination Test for Large Language Models. Presented at: Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL); Dec 6-7, 2023; Singapore p. 314-334. [doi: 10.18653/v1/2023.conll-1.21]
20. Miao J, Thongprayoon C, Suppadungsuk S, Garcia Valencia OA, Cheungpasitporn W. Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications. Medicina (Kaunas) 2024 Mar 8;60(3):445. [doi: 10.3390/medicina60030445] [Medline: 38541171]
21. Oladimeji OO, Oladimeji A, Oladimeji O. Classification models for likelihood prediction of diabetes at early stage using feature selection. ACI 2024 Jun 17;20(3-4):279-286. [doi: 10.1108/ACI-01-2021-0022]
22. Kokori E, Olatunji G, Aderinto N, et al. The role of machine learning algorithms in detection of gestational diabetes; a narrative review of current evidence. Clin Diabetes Endocrinol 2024 Jun 25;10(1):18. [doi: 10.1186/s40842-024-00176-7] [Medline: 38915129]
23. Kumar M, Chen L, Tan K, et al. Population-centric risk prediction modeling for gestational diabetes mellitus: a machine learning approach. Diabetes Res Clin Pract 2022 Mar;185:109237. [doi: 10.1016/j.diabres.2022.109237] [Medline: 35124096]
24. Mennickent D, Rodríguez A, Opazo MC, et al. Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. Front Endocrinol (Lausanne) 2023;14:1130139. [doi: 10.3389/fendo.2023.1130139] [Medline: 37274341]
25. Wu YT, Zhang CJ, Mol BW, et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. The Journal of Clinical Endocrinology & Metabolism 2021 Mar 8;106(3):e1191-e1205. [doi: 10.1210/clinem/dgaa899]
26. Lu HY, Ding X, Hirst JE, et al. Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. IEEE Rev Biomed Eng 2023;17:98-117. [doi: 10.1109/RBME.2023.3242261]
27. Pinto Y, Frishman S, Turjeman S, et al. Gestational diabetes is driven by microbiota-induced inflammation months before diagnosis. Gut 2023 May;72(5):918-928. [doi: 10.1136/gutjnl-2022-328406] [Medline: 36627187]
28. Cubillos G, Monckeberg M, Plaza A, et al. Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy. BMC Pregnancy Childbirth 2023 Jun 23;23(1):469. [doi: 10.1186/s12884-023-05766-4] [Medline: 37353749]
29. Wu J, Zhu J, Qi Y, et al. Medical Graph RAG: evidence-based medical large language model via graph retrieval-augmented generation. : Association for Computational Linguistics Presented at: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1); Jul 27 to Aug 1, 2025; Vienna, Austria p. 28443-28467. [doi: 10.18653/v1/2025.acl-long.1381]
30. Nazi ZA, Peng W. Large language models in healthcare and medical domain: a review. Informatics (MDPI) 2024;11(3):57. [doi: 10.3390/informatics11030057]
31. Ge J, Sun S, Owens J, et al. Development of a liver disease-specific large language model chat interface using retrieval-augmented generation. Hepatology 2024 Nov 1;80(5):1158-1168. [doi: 10.1097/HEP.0000000000000834] [Medline: 38451962]
32. Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. NPJ Ment Health Res 2024 Apr 2;3(1):12. [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
33. Tripathi S, Sukumaran R, Cook TS. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. J Am Med Inform Assoc 2024 May 20;31(6):1436-1440. [doi: 10.1093/jamia/ocad258] [Medline: 38273739]
34. Ong JCL, Jin L, Elangovan K, et al. Development and testing of a novel large language model-based clinical decision support system for medication safety in 12 clinical specialties. arXiv. Preprint posted online on Jan 29, 2024. [doi: 10.48550/arXiv.2402.01741]
35. Semantic Scholar API. Semantic Scholar. 2023. URL: https://www.semanticscholar.org/product/api [accessed 2025-03-18]
36. ChatGPT: optimizing language models for dialogue. OpenAI. 2023. URL: https://openai.com/index/chatgpt/ [accessed 2025-12-14]
37. Trajanoska M, Stojanov R, Trajanov D. Enhancing knowledge graph construction using large language models. arXiv. Preprint posted online on May 8, 2023. [doi: 10.48550/arXiv.2305.04676]
38. Neo4j life sciences and healthcare network. Neo4j. URL: https://neo4j.com/developer/life-sciences-and-healthcare [accessed 2025-09-14]
39. Hu Y, Lei Z, Zhang Z, Pan B, Ling C, Zhao L. GRAG: Graph Retrieval-Augmented Generation. In: Findings of the Association for Computational Linguistics: NAACL 2025: Association for Computational Linguistics; 2025. URL: https://aclanthology.org/2025.findings-naacl.232/ [accessed 2025-09-14] [doi: 10.18653/v1/2025.findings-naacl.232]
40. Claude. 2023. URL: https://www.anthropic.com/claude [accessed 2025-12-14]
41. Labrak Y, Bazoge A, Morin E, Gourraud PA, Rouvier M, Dufour R. BioMistral: a collection of open-source pretrained large language models for medical domains. In: Findings of the Association for Computational Linguistics: ACL 2024: Association for Computational Linguistics; 2024. [doi: 10.18653/v1/2024.findings-acl.348]

XSL•FO
RenderX

42. Ngo NT, Van Nguyen C, Dernoncourt F, Nguyen TH. Comprehensive and practical evaluation of retrieval-augmented generation systems for medical question answering. arXiv. Preprint posted online on Nov 14, 2024. [doi: 10.48550/arXiv.2411.09213]
43. Xiong G, Jin Q, Lu Z, Zhang A. Benchmarking retrieval-augmented generation for medicine. In: Findings of the Association for Computational Linguistics: ACL 2024: Association for Computational Linguistics; 2024:6233-6251. [doi: 10.18653/v1/2024.findings-acl.372]
44. Dobkin R. A man asked ChatGPT how to remove sodium chloride from his diet. It landed him in the hospital. The Independent. 2025 Aug 7. URL: https://www.independent.co.uk/news/health/chatgpt-medical-advice-hospital-b2803992.html [accessed 2025-12-22]
45. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 2024 Sep;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]
46. Ruba F, Nazir A, Evangelista E, Bukhari S, Loft L, Sharma R. GraphRAG-architecture-of-a-local-LLM-for-gestational-diabetes-mellitus. GitHub. 2025. URL: https://github.com/fathima-ruba/GraphRAG-Architecture-of-a-local-LLM-for-Gestational-Diabetes-Mellitus [accessed 2025-12-08]

Abbreviations

AI: artificial intelligence
BLEU: bilingual evaluation understudy
GDM: gestational diabetes mellitus
GenAI: generative artificial intelligence
GP: general practitioner
GraphRAG: graph-based retrieval-augmented generation
KG: knowledge graph
LLM: large language model
ML: machine learning
PoC: proof-of-concept
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAG: retrieval-augmented generation
SML: small language model
T2D: type 2 diabetes

Edited by R Zhao; submitted 23.Apr.2025; peer-reviewed by M Azam, Y Xu, Y Dong; revised version received 14.Oct.2025; accepted 14.Oct.2025; published 05.Jan.2026.

Please cite as:
Evangelista E, Ruba F, Bukhari S, Nazir A, Sharma R
GraphRAG-Enabled Local Large Language Model for Gestational Diabetes Mellitus: Development of a Proof-of-Concept
JMIR Diabetes 2026;11:e76454
URL: https://diabetes.jmir.org/2026/1/e76454
doi:10.2196/76454

© Edmund Evangelista, Fathima Ruba, Salman Bukhari, Amril Nazir, Ravishankar Sharma. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 5.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personalized Type 1 Diabetes Management: Reinforcement Learning–Based Insulin Dosing and Glucose Forecasting

Ernest M Taku[1], BA, MBA; Vibhuti Gupta[2], BS, MS, PhD; Ashutosh Singhal[1], PhD

[1]Department of Biomedical Data Science, School of Applied Computational Sciences, Meharry Medical College, Nashville, TN, United States

[2]Department of Biostatistics & Data Science, School of Public and Population Health, The University of Texas Medical Branch at Galveston, 301 University Boulevard, Galveston, TX, United States

Corresponding Author:
Vibhuti Gupta, BS, MS, PhD
Department of Biostatistics & Data Science, School of Public and Population Health, The University of Texas Medical Branch at Galveston, 301 University Boulevard, Galveston, TX, United States

Abstract

Background: Optimizing insulin dosing and predicting future glucose levels for people with type 1 diabetes is challenging due to the dynamic nature of glucose metabolism. Traditional static insulin regimens fail to adapt to individual variability in diet, physical activity, stress, and metabolic fluctuations, leading to suboptimal glycemic control. Reinforcement learning (RL) offers a promising alternative by enabling personalized, real-time insulin adjustments that improve the balance between hyperglycemia and hypoglycemia.

Objective: This study aims to develop a deep Q-network (DQN)–based RL system that dynamically personalizes insulin dosing recommendations using continuous glucose monitoring data, meal intake, and physical activity levels. By leveraging real-time data, the model adapts to patients' evolving physiological states, enhancing glucose control and patient safety.

Methods: We used the OhioT1DM dataset (2018 and 2020), which includes 8 weeks of continuous glucose measurements, insulin dosing records, and physical activity data for twelve people with type 1 diabetes. The RL agent was designed with a state representation consisting of recent blood glucose levels, insulin doses, and lifestyle factors over a 2-hour window. The 2-hour window was selected based on the known pharmacodynamic profile of rapid-acting insulin (peak action within 90 - 120 min), as well as the typical lag in glycemic response following meals or exercise. This window size captures both recent and delayed physiological effects while balancing data density and model stability. The action space included discrete insulin dose recommendations (eg, 0.5 U, 1 U, and 1.5 U). A reward function incentivized glucose levels within the target range (70 - 180 mg/dL) while penalizing extreme deviations. The DQN model was trained to maximize reward by learning optimal dosing strategies through iterative trial and error.

Results: Performance evaluation was conducted using both qualitative and quantitative metrics. Time-series analysis compared actual and predicted glucose levels, demonstrating effective glucose regulation. The RL model achieved a mean glucose level of 80.06 mg/dL, with a reward score of 10 during evaluation, indicating that most glucose predictions were maintained within the desired clinical range. This suggests the model has learned to regulate blood glucose effectively through adaptive insulin dosing. The root mean square error (12.39 mg/dL) was slightly higher than the mean absolute error (9.85 mg/dL), indicating stable predictions. Additionally, the percentage time in target range was 64.06%, suggesting that the model-maintained glucose within the clinically safe range for a majority of the time.

Conclusions: The DQN-based RL model demonstrated its effectiveness in personalized insulin dosing while minimizing the risk of hypo- and hyperglycemia. This suggests the model has learned to regulate blood glucose effectively through adaptive insulin dosing. This approach represents a significant advancement over conventional methods, offering a scalable and adaptive strategy for real-world diabetes management, along with enhancing clinical trust and transparency through explainability techniques.

(JMIR Diabetes 2026;11:e79195) doi:10.2196/79195

KEYWORDS

personalized insulin dosing; reinforcement learning; deep Q-network; adaptive insulin regimens; machine learning; health care; artificial intelligence

XSL•FO
RenderX

Introduction

Background

According to the National Institute of Diabetes and Digestive and Kidney Diseases, diabetes is a chronic condition affecting more than 38.7 million people in the United States, with approximately 1.6 million diagnosed with type 1 diabetes (T1D) [1]. T1D is an autoimmune condition where the body's immune system mistakenly attacks and destroys the insulin-producing beta cells in the pancreas, resulting in little to no insulin production in the body [2]. Insufficient insulin levels in individuals with T1D can cause hypoglycemia (ie, low blood sugar), hyperglycemia (ie, high blood sugar), and ketoacidosis (ie, ketone development), with potential impacts on vital organs including the heart, kidneys, eyes, and feet. Although the underlying cause of T1D is still unknown, it is widely accepted that both genetic susceptibility and environmental influences contribute significantly to its development [3]. Management of T1D primarily involves regulating blood glucose (BG) levels through insulin therapy, nutritional adjustments, physical activity, and routine glucose monitoring.

Insulin is a hormone the body uses to allow sugar (glucose) to enter cells to produce energy, and it plays a critical role in T1D. Thus, insulin dosing through injections or an insulin pump is the primary treatment to compensate for the body's inability to produce insulin. However, managing diabetes effectively involves precise insulin dosing to maintain optimal BG levels [4]. Due to variability in patients' responses to insulin based on factors such as diet, physical activity, stress, and metabolic fluctuations, the challenge of personalizing insulin dosing is critical for improved patient outcomes. Despite advancements in continuous glucose monitoring (CGM) technologies, insulin dosing remains predominantly reactive and static. These limitations underscore the need for intelligent, adaptive systems capable of personalized treatment.

With the rapid advancement of artificial intelligence (AI) and machine learning (ML) technologies, along with the growing availability of big data in health care, the integration of AI into health care is becoming increasingly feasible [5]. AI's growing role in health care has driven innovations in areas such as image analysis, disease diagnosis and prognosis, clinical decision support, robotic surgery, virtual assistants, and drug target screening [6-9]. However, there are still challenges with the dynamic adaptiveness and explainability of the models. Thus, we leveraged the potential of AI and ML to develop an adaptive and explainable system for personalized insulin dosing for people with T1D.

Several studies [10-15] have examined ML methods for diagnosing and predicting the early onset of type 2 diabetes mellitus (T2DM). Most of these works have applied multiple ML algorithms such as decision trees, support vector machines, random forests, gradient boosting, k-nearest neighbors, neural networks, etc., using various demographics and clinical variables to diagnose and predict diabetes onset. These approaches often rely on static training paradigms and lack the adaptability needed to accommodate dynamic or nonstationary data distributions. Many studies [16-19] have explored ML approaches for BG prediction. These studies have used classical time series models, support vector machines, random forests, and long-short term memory models to predict the hypoglycemic events and forecast BG in near-time horizons. These studies are limited in their ability to provide personalized predictions due to an overreliance on CGM data, without incorporating additional contextual factors such as physical activity, carbohydrate intake, and insulin dosage. Furthermore, they often experience interpretability constraints.

The traditional models often lacked the adaptability required to respond dynamically to changes in a patient's condition. This limitation paved the way for reinforcement learning (RL) models capable of continuous learning and adaptation. RL has attracted interest for its ability to dynamically optimize insulin dosing and predict glucose levels in real time by leveraging data from CGMs. There are some works focused on using RL for dynamic insulin recommendation and BG prediction [20-23]. Early works used Q-learning models [20], followed by more advanced methods such as deep Q-networks (DQNs) [21], actor-critic algorithms [23], and model-predictive control [22]. However, most of these models performed well in simulations; they required extensive manual tuning, making them impractical for long-term use in real-world settings. Thus, these approaches lack integration capabilities with real-world datasets or comprehensive evaluation against supervised models. Model explainability is another limitation in these RL-based methods, which is essential for clinical deployment. Additionally, previous studies frequently rely on synthetic datasets, offer limited explainability, and fail to benchmark RL against traditional machine learning models.

Despite the growing success of RL in health care applications, deploying these models in real-world clinical settings remains a challenge due to lack of interpretability in the models. Gottesman et al [24] discussed the practical hurdles of implementing RL models in health care, such as handling noisy data, model interpretability, and patient safety. Our work directly addresses these issues by applying Shapley additive explanations (SHAP) [25] and local interpretable model-agnostic explanations (LIME) [26] for better model explainability and by thoroughly testing the model on various patient datasets to ensure robustness and safety. Additionally, another study [27] aims to highlight the importance of simulating real-world conditions for RL models to ensure their generalization to unseen patient scenarios. We tackle this challenge by incorporating simulated testing and leveraging CGM data to train and test our RL model.

To address the above-mentioned limitations, we propose a DQN-based RL system that dynamically personalizes insulin dosing recommendations using CGM data, meal intake, and physical activity levels using a real-world dataset from the Ohio T1DM challenge [28], compare the effectiveness of the built model with benchmark ML models, and incorporate explainability using SHAP [25] and LIME [26].

Building on prior DQN-based insulin-dosing research, our paper emphasizes an explainability-first RL framework using the real-world multimodal OhioT1DM dataset. We integrate SHAP and LIME analyses with clinically relevant vignettes (eg, premeal bolus timing and postexercise hypoglycemia mitigation)

XSL•FO
RenderX

to contextualize model behavior. Our safety-aware reward rationale is supported by a compact, systematic sensitivity design—including asymmetry, stability, and insulin on-board traces (IOB)-proxy components across multiple glycemic targets—and by a reproducible data pipeline that harmonizes pump or CGM data with wearable sensor signals. Finally, we implement leakage-safe data splits and report mean (SD) with 95% CIs for TIR or TBR or TAR and error metrics to enhance clinical interpretability and methodological rigor.

Our study uniquely focuses on model explainability—an essential factor for clinical trust that is largely overlooked in existing research. In this study, we applied SHAP to assess feature importance within the DQN model. SHAP values quantify the contribution of each input feature to the model's decisions—for instance, illustrating how "time since last insulin dose" influences insulin recommendations. This explainability perspective is lacking in existing works, which have focused solely on the performance of RL algorithms without examining the rationale behind specific dosing decisions. Another distinguishing aspect of our work is the incorporation of LIME, which is used to explain individual insulin dose recommendations by approximating the DQN model with interpretable surrogate models. For instance, a clinician can interpret a recommended dose by analyzing contributing factors like meal timing or prior glucose patterns. By integrating SHAP and LIME, our work bridges the gap between black-box AI models and the level of interpretability necessary for regulatory approval and clinical implementation.

Objective

We have developed a DQN-based RL system that dynamically personalizes insulin dosing recommendations using CGM data, meal intake, and physical activity levels, and has the capability to predict future BG levels at specified intervals (eg, 30 or 60 min) using historical event data. By leveraging real-time data, the model adapts to patients' evolving physiological states, enhancing glucose control and patient safety. We have compared the effectiveness of our model with the benchmark ML models and incorporated explainability using SHAP and LIME to enhance model understanding. The model's prediction can aid in forecasting episodes of hyperglycemia or hypoglycemia and inform optimal insulin dosing and lifestyle adjustments. Our major contributions in this study are as follows:

1. We have proposed and developed a novel and adaptive RL-based framework for personalized insulin dosing recommendations as well as predicting future BG levels. Our framework consists of three major components:
   a. Deep Q-network: DQN uses a value-based approach where a neural network approximates the Q-value function to decide optimal actions for given states. This is distinct because DQN is better suited for discrete action spaces, which aligns well with insulin dosing (eg, no dose, low dose, medium dose, and high dose).
   b. State space representation: this consists of lag features, rolling averages, and time-based features, offering a richer, temporally aware state representation. It explicitly models time since last insulin dose and time since last meal, emphasizing the physiological delay in glucose-insulin dynamics.
   c. Reward design: the reward function penalizes extreme hypo- and hyperglycemia events (<70 and>250 mg/dL), with positive rewards for glucose levels within the target range (70 - 180 mg/dL). Our approach introduces adaptive penalties, which vary with the severity of glucose deviation, potentially improving safety margins.
2. We have performed extensive evaluation of our RL-based framework performance with various metrics and benchmarked the performance with the state-of-the-art long short-term memory (LSTM) model.
3. We have incorporated explainability into the built RL-based model by integrating SHAP and LIME methods to assess feature importance and explain individual insulin dosing recommendations.

Methods

Ethical Considerations

The OhioT1DM dataset is a deidentified dataset requested from the authors through a data use agreement. No institutional review board review or approval is required because the data are completely deidentified.

Dataset and Preprocessing

In this study, we used the OhioT1DM dataset from 2018 and 2020, provided by the Blood Glucose Level Prediction challenge [28,29]. The dataset was generated by monitoring 12 individuals with T1D over an 8-week period, capturing a range of BG-related data. It includes CGM readings recorded every five minutes, insulin delivery data from insulin pumps, and self-reported events such as meals, work, sleep, psychological stress, and physical activity, all logged via a smartphone app. Additionally, physical activity was tracked using a sensor band. The first cohort, consisting of 6 individuals, wore Basis Peak fitness bands; the dataset contains 5-minute aggregated measurements of heart rate, galvanic skin response (GSR), skin temperature, air temperature, and step count [28]. The second cohort, also comprising 6 individuals, wore the Empatica Embrace; the dataset provides 1-minute aggregated measurements of GSR, skin temperature, and acceleration magnitude [28]. Notably, meal and insulin data are represented as discrete user-entered values rather than continuous series like carbohydrate intake or insulin on board. The detailed data description is shown in Table 1.

XSL•FO
RenderX

Table . Dataset description.

Attributes	Description	Unit
timestamp	Date and time of the event	DateTime
glucose_level	Blood glucose level at a specific time	mg/dL
insulin_dose	Administered insulin dose during the event	Units (U)
carbs	Carbohydrate intake associated with a meal	grams (g)
meal_type	Type of meal (eg, breakfast, lunch, and snack)	__[a]
exercise_intensity	Intensity of physical activity	Intensity Level (1-5)
exercise_duration	Duration of the exercise	minutes
bolus_dose	Insulin dose delivered as a bolus	Units (U)
heart_rate	Heart rate during the event	beats per minute (bpm)
sleep_quality	Quality of sleep	Percent (%)

[a]Not applicable.

The OhioT1DM dataset consists of participant demographics: gender distribution (6 male, 6 female), age ranges (20 - 80 y), and device use. All participants were on insulin pump therapy, using Medtronic 530G or 630G models with Medtronic Enlite CGM sensors. Both basal and bolus insulin data are included. While BMI and diabetes duration are not reported in the public dataset, all participants were experienced pump users, and therapy modalities were consistent across the cohort [28]. The detailed description is shown in Table S2 in Multimedia Appendix 1.

The raw dataset is in XML format. We converted the XML files into CSV format first, performed preprocessing steps such as handling missing data and normalizing features, and generated engineered features, including lag, rolling, and time-based attributes, for further analysis. We extracted relevant data such as glucose levels, insulin doses (from bolus events), meals (carbs intake), exercise, and other factors. The key characteristics of the extracted data are shown in Table 1. The extracted data consists of 1,191,753 records with the attributes as shown in Table 1.

The time is aligned to 5 minutes, and Empatica 1-minute channels are down-sampled (mean or median) to 5 minutes before fusion. Missing CGM gaps are imputed using backward or forward fill up to short horizon (<=15 min), and longer gaps are excluded and not being considered for learning or evaluation. The sensor outliers (eg, CGM<40 mg/dL or >400 mg/dL, nonphysiologic spikes) are clipped and flagged. The duplicate or overlapping basal and multipart bolus events (square or dual) are resolved to continuous IOB traces.

Time series and distribution plots were used to identify trends, as illustrated in the figures. Figure 1A and B shows the distribution of glucose levels overall and within a specific range, respectively. As shown in Figure 1A, there is a highly skewed distribution of glucose levels where a significant number of entries have glucose levels close to 0, which could indicate periods of hypoglycemia or potentially erroneous readings. Figure 1B shows the distribution of glucose levels ranging from 50 to 400, with an average level of 150. There is another peak around 100, as shown in Figure 1B, which is closer to typical blood sugar levels but still slightly on the lower side of normal. There are fewer instances of high glucose levels, suggesting fewer episodes of high blood sugar in the dataset.

Figure 2A and B illustrate the glucose levels over time for 2 sample patients from the dataset. Given that glucose levels are timestamped, we analyze how these levels change over time within individual patients and across different days or times of day. We can observe a general downward trend, with minor fluctuations. This suggests the potential usefulness of considering past values (lag features) when predicting future glucose levels. The descriptive statistics for all 12 participants' data are in Table S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. (A) Distribution of glucose levels (left) and (B) distribution of glucose levels in specific range (right).

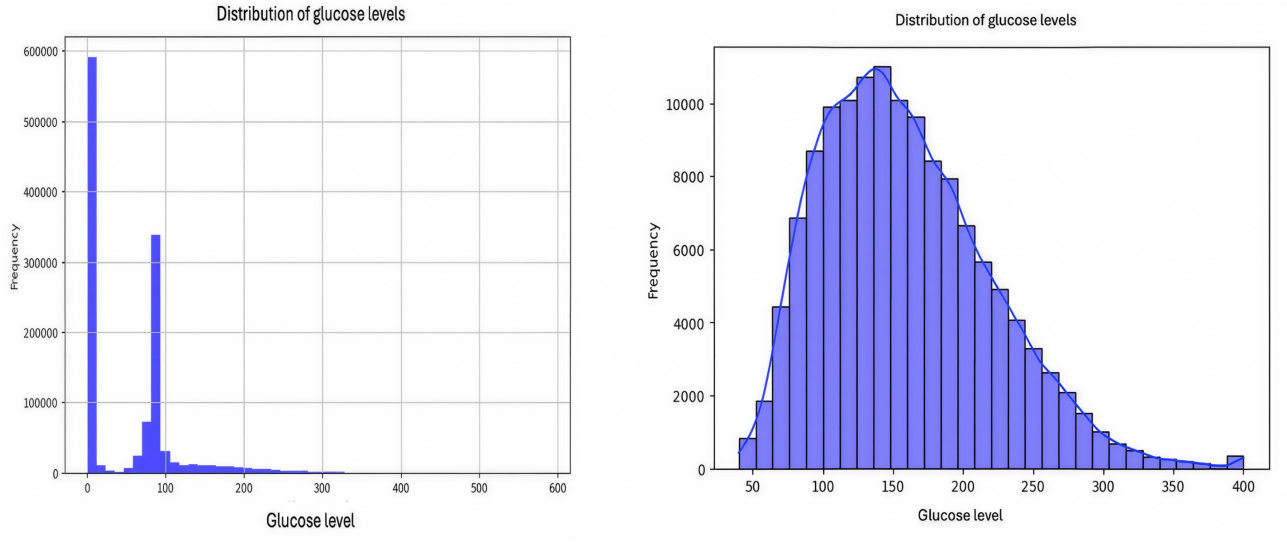
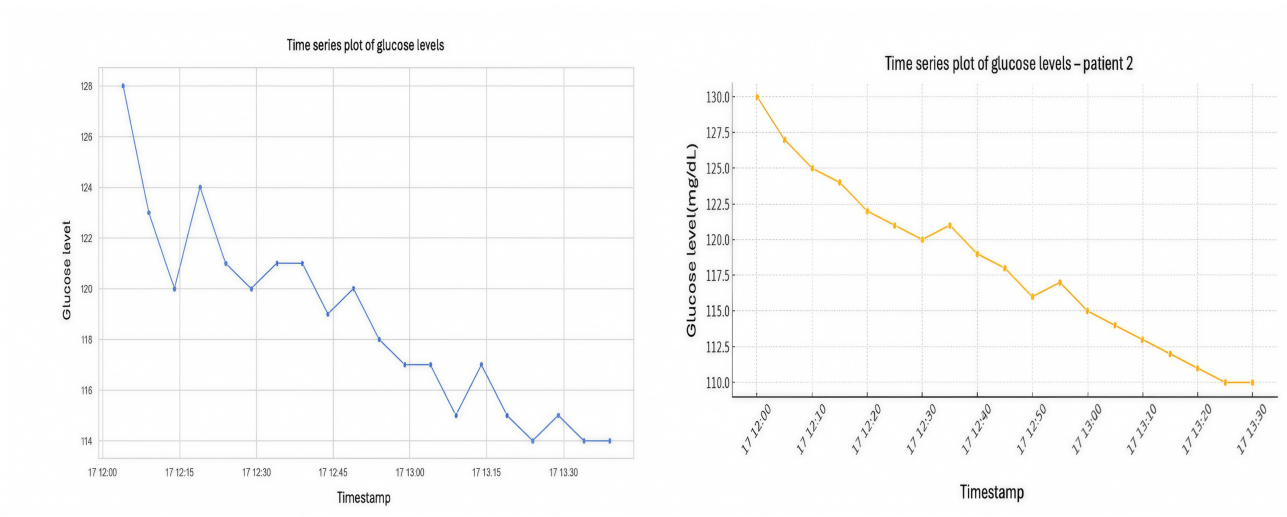


Figure 2. (A) Glucose levels variations for patient 1 (left) and (B) glucose levels variations for patient 2 (right).



Model Development

RL [30] focuses on the task of deriving a policy that maps states to actions in a way that maximizes cumulative reward. These problems are inherently closed-loop, as the agent's actions directly affect subsequent environmental states and observations. Unlike in supervised learning, the agent receives no explicit instruction on which actions to take; it must learn the optimal actions through interaction with the environment, identifying those that maximize reward via trial and error. RL has been successfully applied across various scientific domains, including robotics and control systems [31], manufacturing, and combinatorial search tasks like those found in computer games [32,33]. In the health care domain, RL has leveraged historical medical data, such as medical images and treatment regimens, for tasks including cancer prediction, diagnosis, and prognosis [24,34,35].

RL systems have multiple components: an agent, a policy, a reward signal, a value function, and optionally a model for the environment [30]. In RL, the goal is for an agent to learn a policy that maximizes a cumulative reward through interacting with an environment. A policy represents a mapping from states to actions that dictates the agent's behavior at a specific point in time. It aligns with the concept of stimulus–response associations in psychology, where the term 'stimulus' encompasses both external inputs and internally generated cues within the organism [30]. A reward signal defines the goal of a problem. At each time step, the environment provides the RL agent with a numerical reward. The agent's primary objective is to maximize the cumulative reward it obtains over time [30]. While the reward signal reflects the immediate desirability of a given outcome, the value function captures long-term benefit. Specifically, the value of a state represents the expected cumulative reward an agent can obtain in the future, beginning from that state [30]. The fourth and final component present in some RL systems is a model of the environment. This model simulates the environment's dynamics or, more broadly, enables the agent to make predictions about future environmental responses [30].

We trained a DQN agent to maximize time in target glucose range. States included glucose, insulin, meals, and exercise. Actions were discrete insulin doses. Reward function penalized values outside 70 - 180 mg/dL. The core RL equation is the

XSL•FO
RenderX

Bellman equation, which forms the basis for algorithms such as Q-learning and DQN.

The Bellman Equation for Q-Learning:

$$(1) Q(s,a)=r+\gamma \max_{a'} Q(s',a')$$

- Q (s, a) is the Q-value (the expected future reward) for taking action *a* in state *s*.
- r is the immediate reward received after taking action *a* in state *s*.
- γ is the discount factor (typically between 0 and 1), which determines how much future rewards are worth compared to immediate rewards.
- $\max_{a'} Q(s', a')$ is the maximum future Q-value (the best future reward the agent can achieve) for the next state *s′* and all possible actions *a′*.

The Bellman equation details with respect to the problem are described below:

1. State (s): in the context of diabetes management, the state can be represented as a combination of features such as the current glucose level, insulin dose, meal intake, exercise, etc
2. Action (a): the action refers to the insulin dose recommendation or adjustment (eg, deciding how much insulin to administer at the current state).
3. Reward (r): the reward is a feedback signal to indicate the success of the agent's action. In diabetes management, the reward might penalize for hypo- or hyperglycemia events and give positive rewards for keeping glucose levels within a healthy range.
4. Discount factor (γ): the discount factor controls the balance between prioritizing immediate rewards (eg, maintaining glucose levels right now) and future rewards (eg, preventing long-term health issues caused by poor glucose control).
5. Max future Q-value: the agent estimates the maximum future reward it can achieve by taking the best action in the next state. This is used to help the agent choose actions that will not only lead to immediate benefits but also to longer-term gains.

Reinforcement Learning Model Selection

The following steps describe the reinforcement learning model selection and training process:

1. Build and train a reinforcement learning model: develop a DQN model to optimize insulin dosing based on extracted parameters such as glucose levels. Details for the DQN model are in the section below.
2. Evaluate and optimize the RL model: after training, we evaluate the RL model's performance using both datasets [28,29], optimize it, and detail the evaluation metrics.
3. Establish state representation: BG levels, insulin doses, meal intake, and physical activity over a recent time window (eg, 2 h).
4. · Reward function: a penalized reward system that reduces points for glucose levels outside the target range, with larger penalties for extreme hypo- or hyperglycemia events.
5. Implement the training loop for the DQN agent to interact with our diabetes management environment by setting up the environment and the agent, then running through multiple episodes to allow the agent to learn optimal actions based on the given state.
6. We set up the DQN agent and its learning mechanisms, and we train this agent using the simulation environment by repeatedly interacting with it (using the step function) and applying the replay function to learn from past actions.

DQN Rule

In the context of DQNs, the Bellman equation is approximated using neural networks to learn the Q-value function. The update rule is:

$$(2) Q(s,a) \leftarrow Q(s,a)+\alpha(r+\gamma \max Q(s',a')-Q(s,a))a'$$

- α is the learning rate that controls how much the Q-values are updated at each step.
- The expression $r+\gamma \max_{a'} Q(s',a')$ is the target Q-value, and the difference $r+\gamma \max_{a'} Q(s',a')-Q(s,a)$ is the temporal difference error.

Details With Respect to the Problem

The following key concepts are central to understanding the problem formulation in reinforcement learning–based diabetes management systems:

1. Policy: the policy defines the agent's behavior — which actions to take in different states. The agent seeks to learn an optimal policy that maximizes the total reward.
2. Exploration versus exploitation: to learn the best policy, the agent must explore different actions (exploration) while also choosing actions it believes will give the best reward based on its current knowledge (exploitation).

In the diabetes management system, the agent continuously learns to recommend insulin doses based on real-time data such as glucose levels, insulin history, and meal intake, balancing immediate glucose control with long-term health management.

Feature Engineering

We have extracted below features after data analysis:

1. Lag features: these capture the prior values of glucose levels. Given the data collection frequency of every 5 minutes, we have created lag features representing glucose levels from 30 to 60 minutes prior.
2. Rolling window features: these features consist of rolling averages to smooth out fluctuations and capture trends over time, containing rolling means and standard deviations over various windows, such as 30 minutes and 60 minutes.
3. Time-based features: since physiological responses can vary throughout the day, including features such as the hour of the day, might capture these variations effectively. We extracted time-based features from the timestamp such as the hour of the day.

Evaluation Metrics

The performance was evaluated quantitatively using standard metrics such as mean absolute error (MAE), root mean square error (RMSE), and time-in-range (TIR). MAE [36] and RMSE [37] are used to assess the accuracy of predictive models for glucose forecasting; however, TIR [38] represents the percentage

XSL•FO
RenderX

of time a person's glucose levels remain within the target range, typically 70 - 180 mg/dL for most adults with diabetes.

The MAE [36] is the average of the absolute errors (ie, the difference between the actual and predicted glucose values).

$$(3)\text{MAE}=1n\sum i=1n|ypredictedi-yactuali|$$

It can be represented, as shown in Equation 3, where n is the number of test instances. A lower MAE value leads to a better model.

RMSE [37] is the square root of the average of squared errors (ie, the difference between the actual and predicted popularity values). RMSE can be represented as shown in Equation 4, where n is the number of test instances. A lower RMSE value leads to a better model.

$$(4)\text{RMSE}= 1n \sum i=1n(ypredictedi- yactuali)2$$

TIR is represented in Equation 5, where higher TIR is associated with better glucose control as compared to lower TIR values.

$$(5)\text{TIR}=(\text{Total count of glucose reading within range}\text{Total number of reading})\times 100$$

Hypothesis

We expect the RL-based model to outperform baseline approaches by increasing the percentage of time glucose levels remain in the target range while minimizing hypo- and hyperglycemia events.

Simulation Outcome

State

The following describes the simulation state:

- Glucose level: 80.06 mg/dL, which falls within the target range (70 - 180 mg/dL). This is a healthy and realistic level for a patient managing diabetes, indicating that the simulation now better reflects the models' physiological responses to insulin.
- Time since last meal: 1 hour, which is a typical scenario and can influence immediate subsequent readings.
- Time since last dose: 1 hour, reflecting recent insulin activity, which might be stabilizing BG.
- Insulin type: generic bolus (mealtime) and basal (background) insulin doses.
- Time of day: 12 (noon), a common time for a meal which might coincide with a postprandial glucose reading.

Reward

The maximum positive reward is 10, reflecting optimal glucose management in the simulation. This reward confirms that the glucose level is within the desired range, and the model's reward system is functioning as intended to encourage similar outcomes. The adjusted parameters helped align the simulation more closely with realistic diabetes management scenarios. The positive reward outcome encourages the model to replicate or aim for similar decisions under comparable circumstances, reinforcing good management practices. We implemented a conservative, interpretable reward function prioritizing glycemic safety, consisting of in-range bonuses (70 - 180 mg/dL) and out-of-range penalties, with an asymmetric variant that assigns stronger penalties to hypoglycemia than to hyperglycemia. To evaluate robustness, we conducted a structured sensitivity analysis across (1) alternative glycemic thresholds (70 - 180, 80 - 160, and 70 - 140 mg/dL) and (2) incremental reward components, including the baseline formulation, asymmetric penalties, a stability penalty on $|\Delta G/\Delta t|$ (ie, penalizes rapid glucose changes reducing glycemic variability where ΔG represents the glucose change with respect to time t), and an IOB proxy penalty to discourage insulin stacking. Across conditions, we summarize TIR or TBR or TAR and MAE or RMSE, showing that safety-focused asymmetry consistently reduces TBR with minimal impact on TIR. These findings support the use of a simple yet clinically aligned reward structure for this initial study.

Results

Experimental Setup

The models were trained for 10 episodes using simulated environments initialized with realistic patient data extracted from the OhioT1DM dataset. The model architecture consisted of a 3-layer fully connected neural network with ReLU activations, Adam optimizer (learning rate 0.001), epsilon-greedy exploration (ε decay from 1.0 to 0.1 over episodes), and a replay buffer size of 5000. We have used time-blocked split per subject consisting of weeks 1 - 6 as the training set, week 7 as validation, and week 8 as the test set. We have used TensorFlow 2.12 for neural network implementation of DQN agent, OpenAI Gym for simulating the agent's interaction with glucose trajectories and insulin dosing responses, Pandas and NumPy for data preprocessing and state-space formulation, and Matplotlib and seaborn for data visualization. The environment represented each state as a vector composed of time-series values over a 2-hour sliding window of glucose levels, meal carbohydrate intake, physical activity intensity or duration, and prior insulin doses. The experiments were conducted on a local machine with Apple M1 Pro 10-core CPU and 32 GB RAM. Training durations were intentionally limited to 10 episodes due to computational constraints, which balances exploratory learning without excessive runtime. The learning curves showed stabilization by episode 8, with negligible performance improvement thereafter (ΔTIR<0.5%). The learning curves demonstrated an early plateau in episodic reward and TIR, indicating initial stabilization but not the full convergence typically expected in deep RL. Future work will extend training to at least 300 episodes, incorporate soft target updates (τ=.005), maintain a replay buffer of at least 100 k transitions, decay exploration to ε=.05, and use leave-one-subject-out (LOSO) cross-validation. Each episode consisted of ~1000 - 2000 state transitions (steps), depending on the patient's data length and time granularity. To ensure reproducibility, random seeds were set for NumPy and TensorFlow. Model checkpoints were stored and reloaded for inference and evaluation. Performance metrics such as MAE, RMSE, and TIR% were computed on unseen test data segmented from each patient's profile. Reward shaping and epsilon decay were used for policy exploration. Our study provides valuable insights into the performance of the DQN agent in managing glucose levels through insulin dosing.

XSL•FO
RenderX

Integration With DQN Algorithm for Training

The simulation environment produces realistic and consistent results when we integrate with the DQN model for training. We initialize the DQN Agent by setting up the neural network model that will learn the Q-values. Then simulate interactions while running episodes where the agent interacts with the environment, makes decisions based on its current policy, observes rewards, and updates its policy accordingly. We then proceeded with the training loop where for each episode, reset the environment. For each time step within the episode, choose an action from the DQN agent, observe the new state and reward, and store this experience. And periodically update the DQN agent's neural network by replaying a batch of experiences.

Model Performance

We compared the glucose prediction performances of two models (ie, DQN and LSTM) in this study. We observed that DQN achieves a slightly lower RMSE (12.39) as compared to LSTM (12.87), indicating it makes fewer large-scale prediction errors. RMSE penalizes larger errors more heavily, so this suggests that DQN is better at avoiding extreme outliers in glucose level prediction. Moreover, DQN is more stable and robust in maintaining glucose predictions closer to the true values, particularly in high-variance regions. LSTM achieves a significantly lower MAE (3.69), suggesting its day-to-day average predictions are closer to the actual glucose levels than DQN's. However, MAE does not penalize outliers as strongly as RMSE. So, while LSTM is generally more accurate on average, it is more vulnerable to large mistakes, as reflected by its higher RMSE.

The DQN model (64.06%) outperforms LSTM (62.10%) slightly for TIR. This is critical in real-world diabetes management, where maintaining glucose in a safe range directly correlates with reduced risk of complications. Since DQN is an RL-based model, reward is a key internal measure of how well it is optimizing glucose control based on its policy. A higher average reward for DQN (39.09) indicates it is more effective in learning strategies that lead to favorable outcomes (ie, TIR maintenance and avoiding hypo- or hyperglycemia). LSTM is not a reinforcement model, so its reward is derived post hoc and may not reflect learning behavior but rather performance fitting (Figure 3).

Figure 3. Model performance comparison for deep-Q-network with LSTM. DQN: deep-Q-network; LSTM: long short-term memory; MAE: mean absolute error; RMSE: root mean square error; TIR: time-in range.

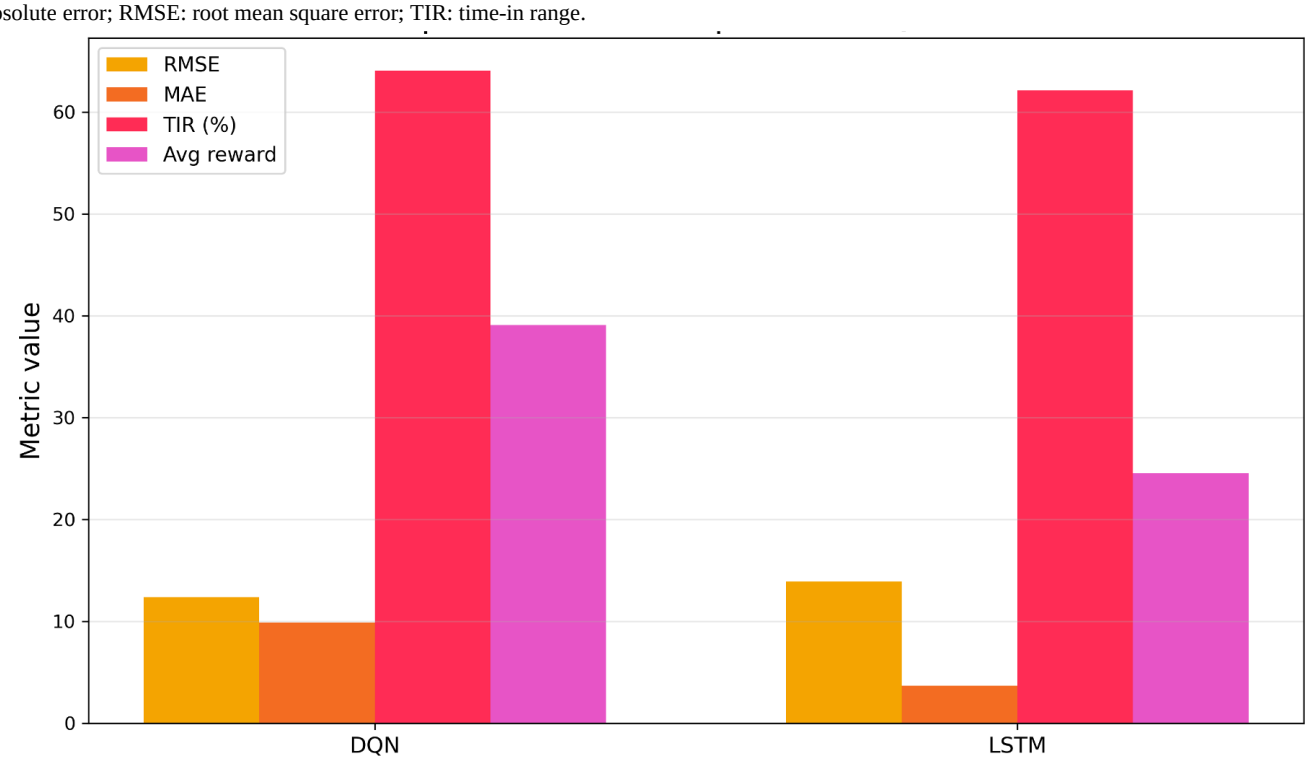


Table 2 summarizes the comparative insights for the model performance using all evaluation metrics between DQN and LSTM models.

XSL•FO
RenderX

Table . Summary of comparative insights.

Dimension	Better model	Reason
RMSE[a]	DQN[b]	Fewer large-scale prediction errors
MAE[c]	LSTM[d]	Closer average prediction to ground truth
TIR[e] (%)	DQN	Maintains glucose in safe range longer
Reward	DQN	Optimizes policy to reinforce good outcomes

[a]RMSE: root mean square error.
[b]DQN: deep-Q-network.
[c]MAE: mean absolute error.
[d]LSTM: long short-term memory.
[e]TIR: time in range.

Table 3 summarizes the statistical analysis of model performance between DQN and LSTM models. On average, the RMSE across models is about 13.1 mg/dL. The relatively small SD shows little variability between DQN (12.39) and LSTM (13.87). The wide CI reflects the small sample size (only 2 models), so additional baselines would stabilize the estimate. The MAE varied more strongly between models (DQN=9.85 vs LSTM=3.69), leading to a large SD and an unrealistic CI range. This shows that comparing only 2 models provides limited inferential strength. LSTM had notably lower MAE, which suggests better average prediction closeness, though this must be contextualized with other metrics. Both models kept patients in the clinical target range (70 - 180 mg/dL)~63% of the time. The small SD indicates stable performance between DQN (64.06%) and LSTM (62.10%). Still, the CI is wide because of limited sample points; more repeated trials are needed to confirm robustness. The reward variance is high due to differences in RL design. DQN achieved a much higher average reward (39.09) than LSTM (24.54). The wide CI reflects the volatility of the RL process and the insufficient number of episodes (only 10, per earlier feedback).

Table . Summary statistics of model performance.

Metric	Mean (SD)	95% CI
RMSE[a]	13.13 (1.05)	(3.73-22.53)
MAE[b]	6.77 (4.36)	(–32.37 to 45.91)
TIR[c] (%)	63.08 (1.39)	(50.63%-75.53%)
Reward	31.82 (10.29)	(–60.62 to 124.25)

[a]RMSE: root mean square error.
[b]MAE: mean absolute error.
[c]TIR: time in range.

Overall, RMSE and TIR are stable across models, suggesting both frameworks maintain reasonable glucose control. The MAE favors LSTM, but this may be influenced by LSTM's smoother short-term predictions, whereas DQN optimizes long-term control. The reward strongly favors DQN, aligning with its reinforcement design, but instability remains due to a few training episodes. Statistical confidence is weak due to a very small sample size (n=2 models); adding baselines (eg, autoregressive integrated moving average, random forest, and simple linear predictors) would allow meaningful variance testing (ANOVA).

Explainability Analysis of RL Model

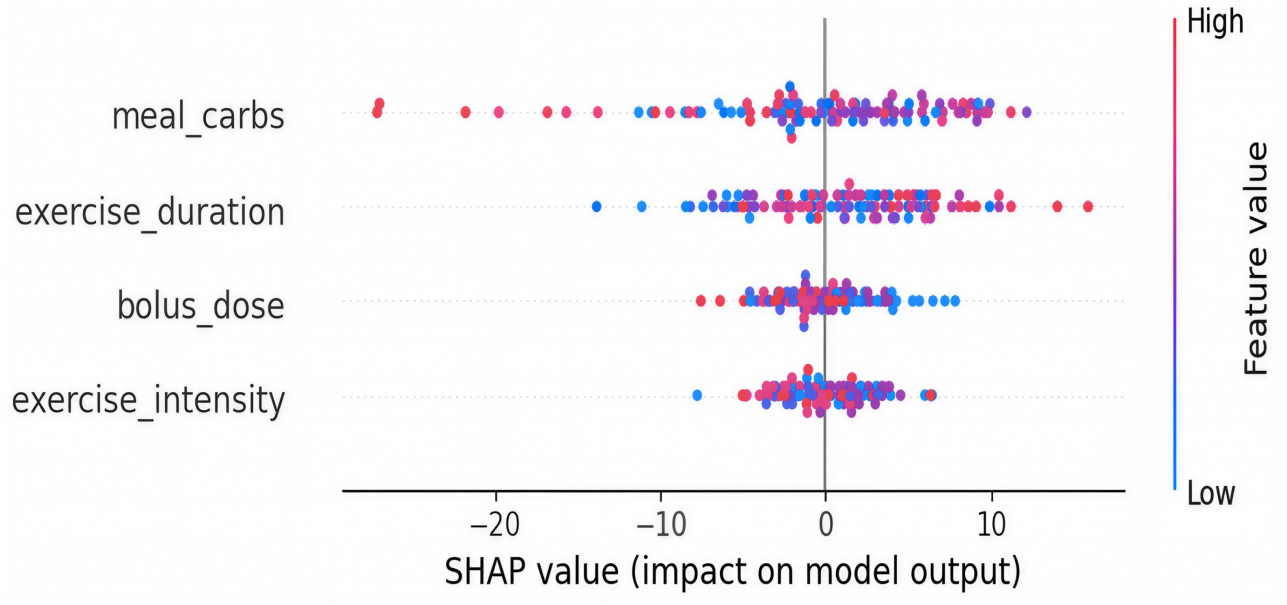
The RL model provides a clinically aligned strategy for insulin management. It not only predicts future glucose values but acts upon them in a self-improving feedback loop, a defining characteristic of RL. We implemented an explainability approach in the model using techniques such as SHAP and LIME, so that clinicians can understand these decisions, aiding interpretability and safety validation in future real-world trials.

SHAP Summary Plot

The SHAP [25] summary plot provides a global view of the feature importance and direction of impact on the model's glucose level predictions. We have applied the SHAP method to improve the interpretability of our RL model. The feature values are represented in color from low (blue) to high (red), and their impact (SHAP value) is plotted on the x-axis as shown in Figure 4.

XSL•FO
RenderX

Figure 4. SHAP plot. SHAP: Shapley additive explanations.



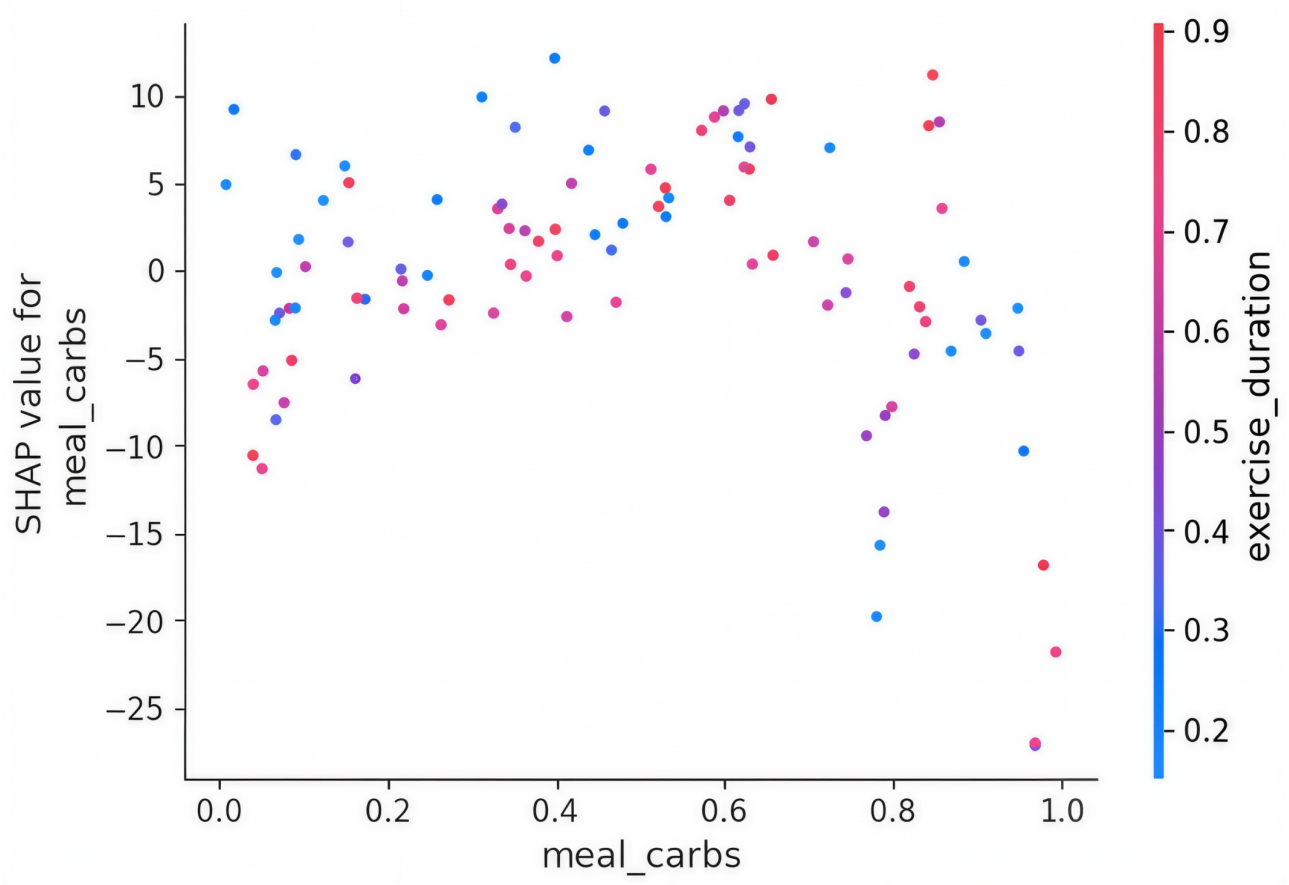
As shown in Figure 4, “meal_carbs” shows the highest SHAP impact range, varying from approximately −25 to +10. High meal_carbs (red) significantly increase predicted glucose. “exercise_duration” ranges in SHAP values from about −10 to +10, indicating that longer durations can both decrease or modestly increase glucose depending on context. “bolus_dose” mostly impacts glucose predictions negatively, ranging from −10 to +5 SHAP units, where higher doses (red) tend to lower glucose predictions. “exercise_intensity” exhibits mostly negative SHAP values clustered between −7 and 0, reflecting a minor downward pressure on glucose levels with increased intensity.

SHAP Dependence and Interaction Effect of Meal Carbs Feature With Exercise Duration

Figure 5 shows the interaction plot of features meal carbs with exercise duration and their impact on the SHAP values. This dependence plot explores the impact of “meal_carbs” on the SHAP values, which reflects its influence on glucose predictions. The plot also colors each point by “exercise_duration” to show interaction effects, as shown in Figure 5.

XSL•FO
RenderX

Figure 5. SHAP interaction plot. SHAP: Shapley additive explanations.



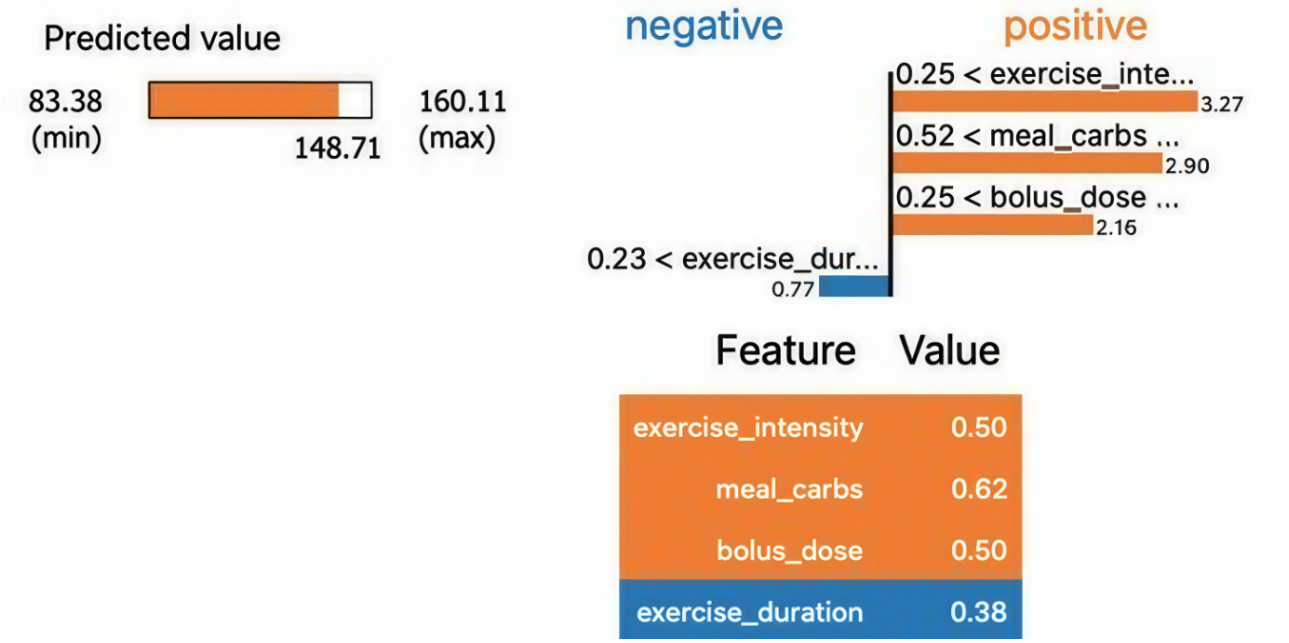
When "meal_carbs" is normalized below 0.2, SHAP values mostly fall between −5 and +5, showing minimal impact. As "meal_carbs" increases to the range of 0.4 - 0.6, SHAP values often rise above +5, with some reaching ~+10. At very high "meal_carbs" values (near 1.0), SHAP values can reach as low as −25 and as high as +10. Points with lower "exercise_duration" (more pink or red) cluster with higher positive SHAP values, indicating increased glucose risk when carbs are high and exercise is low.

LIME Explanation Plot for Individual Predictions

The LIME plot presents the local feature contribution to a single glucose level prediction made by the model. Figure 6 shows the LIME plot to explain the individual features' contribution in the glucose prediction. The positive and negative bars show how much each feature pushed the predicted value up or down for this instance.

XSL•FO
RenderX

Figure 6. Local interpretable model-agnostic explanations plot.



As shown in Figure 6, we observed predicted glucose value: ~148.71 mg/dL (within the model's learned bounds of 83.38 to 160.11 mg/dL). The "meal_carbs" feature at 0.62 contributed +8.27 units to glucose prediction while "*bolus_dose*" at 0.50 contributed +2.90 units. This might indicate a mismatch or delay in insulin effect. The "exercise_intensity" feature at 0.50 contributed +2.27 units while "exercise_duration" feature at 0.38 contributed −0.77 units, helping to modestly reduce predicted glucose.

Clinical Interpretation and Implications

Our model explainability results uncover how key variables influence BG predictions by combining global (SHAP) and local (LIME) interpretability methods. The key takeaways include: high meal carbohydrate intake is a dominant factor raising glucose, especially when not balanced by insulin or physical activity. Insulin dosing ("bolus_dose") generally reduces glucose but can be insufficient if the meal size is high or if timing is off. Exercise, particularly duration, helps buffer glucose spikes and improves prediction outcomes. Personalized decision-making should consider the interaction between carbs, insulin, and exercise—simple rule-based systems may overlook these subtleties. This reinforces the need for dynamic, explainable AI systems in diabetes management to tailor recommendations based on a full patient context.

We ground the interpretability in patient-facing and clinician-facing scenarios, using concrete values we already computed: For example (LIME, single decision): predicted glucose=148.7 mg/dL; features contributing upward: "meal_carbs=0.62" (+8.27 SHAP or LIME units), "bolus:_dose=0.50" (+2.90 units; likely timing mismatch vs IOB), "exercise_intensity=0.50" (+2.27 units), and "exercise_duration=0.38" (−0.77 units) lowering risk. The clinical actionability example is:

1. Premeal guidance: if SHAP shows carbs consistently the top driver (+6 - 10 units) and low IOB, suggest pre-bolus timing (eg, 15 - 20 min) or carb ratio adjustment for that period of day.
2. Postexercise hypoglycemia risk: if SHAP highlights high exercise_duration with negative contributions and declining rate-of-change, the system can warn: reduce correction bolus or add carbs to avoid late-onset lows.
3. Overnight stability: if basal-driven negative SHAP at night with frequent lows, suggest basal down-titration or tighter overnight safety constraints in the policy.

Discussion

Principal Findings

In summary, our results suggest that DQN outperformed static dosing models and matched supervised models such as LSTM in RMSE while offering dynamic control. Explainability analysis revealed critical decision factors. An RMSE of 12.39 mg/dL is also fairly low but slightly higher than the MAE, suggesting that there might be occasional predictions with larger errors. Despite this, the model's overall predictive accuracy is good. Percentage time in target range (64.06%). This metric indicates that the glucose levels are within the clinically safe range (70 - 180 mg/dL) about 64.06% of the time. Although this represents the majority of the time, there remains room for improvement, as optimal diabetes management typically aims for a higher percentage within this range to minimize the risk of complications from high or low BG levels. An average reward of 39.09 indicates that the agent generally performs well under the reward structure we have defined, which presumably rewards the agent for maintaining glucose levels within the target range and penalizes it for deviations.

Clinical Relevance

Explainable RL can improve clinician trust, facilitate regulatory approval, and ensure patient safety. Adaptive insulin recommendations reduce the risk of adverse glycemic events. General observations show that the majority of values lie below

XSL•FO
RenderX

200 mg/dL, which is generally considered within a manageable range for people with diabetes but includes many readings below 70 mg/dL, which are hypoglycemic. The upward trend in the data visualization suggests an increase in average glucose levels over time, though this could also reflect variabilities in patient behavior or treatment efficacy. The output from the simulation showed a glucose level of approximately 80.06 mg/dL and a reward of 10, suggesting that the adjustments to the insulin sensitivity and decay parameters have significantly improved the realism and functionality of the environment. DQN is superior in handling real-time decisions and dynamic conditions, making it more suitable for adaptive insulin dosing in personalized care. LSTM could be valuable in applications where predicting general trends is sufficient (eg, retrospective analysis and forecast dashboards). Integrating hybrid architectures (eg, DQN for policy and LSTM for predictive enhancement) may yield optimal results. Ultimately, DQN's better TIR and reward profile signal its strength in closed-loop, autonomous glucose management systems.

Limitations and Future Work

Although the results are promising, our work has certain limitations. First, the data were limited to 12 participants which limits the applicability of our work to real-time deployment. However, it provides enough evidence of using RL-based models for glucose prediction and insulin dosing recommendations. A larger dataset will help to validate the models better. The future work includes hybrid DQN-LSTM ensembles, generalization to Type 2 diabetes. We aim to apply continuous enhancement of the models to optimize performance by focusing on key components such as:

1. Data quality and quantity: ensuring high-quality and comprehensive data can help improve model accuracy. This includes a detailed recording of insulin doses, meals, exercise, glucose levels, etc.
2. Feature engineering: explore different features or combinations that might improve model predictions, such as time of day, preceding meal types, or exercise intensity.
3. Model tuning: for LSTM and DQN, parameter tuning could optimize performance. This involves adjusting learning rate, number of trees, depth of trees, etc.
4. Ensemble techniques: combining predictions from different models (eg, an ensemble of LSTM and DQN) might leverage strengths and mitigate individual weaknesses.
5. Incremental training: for LSTM and DQN, consider using an incremental training approach to continually update the models as new data becomes available, which might help in adapting to changes in patients' lifestyle or insulin sensitivity.

Conclusions

We demonstrate that a DQN-based RL system can effectively personalize insulin dosing in T1D. The system achieved strong predictive performance, maintained glucose within safe ranges, and enhanced interpretability through SHAP and LIME. This work paves the way for clinically integrated, AI-assisted diabetes care systems. The analysis shows that while the insulin type does not lead to significant differences in glucose control, individual management strategies, including timing and dosage adjustments, are critical. The variability seen both within and across patients suggests the need for personalized diabetes management plans, closely monitored by health care providers. The frequent fluctuations and the presence of high and low extremes indicate a need for a reinforcement learning model that can dynamically adjust insulin dosages based on CGM data to better manage and stabilize glucose levels. In reinforcement learning models such as DQN, the reward function plays a key role. If the reward function is not designed correctly, the model might not learn the correct associations between actions and outcomes (predictions). A poor reward signal could result in the model always predicting values close to zero.

Funding

The research reported in this article was supported by AIM-AHEAD Coordinating Center, award number OTA-21-017, and was, in part, funded by the National Institutes of Health (NIH) Agreement No. 1OT2OD032581-02-836. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

Data Availability

Prior to March 17, 2026, access to the dataset required an executed Data Use Agreement (DUA). The dataset is now available upon request through the website [39]. The dataset curator has no objection to the publication of this article, and the previous DUA requirement has been waived.

Authors' Contributions

EMT contributed to the investigation, methodology, data analysis, visualization, and writing the original draft of the manuscript. VG contributed to data curation, resources, supervision, as well as reviewing and editing the manuscript. AS contributed to co-supervision, conceptual framing, data interpretation, clinical context, manuscript restructuring, and critical revisions. All authors reviewed and approved the final version.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Descriptive statistics of glucose levels and cohort characteristics.
[DOCX File, 20 KB - diabetes_v11i1e79195_app1.docx]

References

1. Diabetes statistics. National Institute of Diabetes and Digestive and Kidney Diseases. URL: https://www.niddk.nih.gov/health-information/health-statistics/diabetes-statistics [accessed 2025-04-23]
2. Type 1 diabetes. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011 [accessed 2025-04-23]
3. Type 1 diabetes. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/diabetes/type-1-diabetes [accessed 2025-04-23]
4. Hirsch IB. Type 1 diabetes mellitus and the use of flexible insulin regimens. Am Fam Physician 1999 Nov 15;60(8):2343-2352. [Medline: 10593324]
5. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023 Jan 13;23(1):7. [doi: 10.1186/s12911-023-02103-9] [Medline: 36639799]
6. Keskinbora KH. Medical ethics considerations on artificial intelligence. J Clin Neurosci 2019 Jun;64:277-282. [doi: 10.1016/j.jocn.2019.03.001] [Medline: 30878282]
7. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019 Jan;25(1):70-74. [doi: 10.1038/s41591-018-0240-2] [Medline: 30617318]
8. Álvarez-Machancoses Ó, Fernández-Martínez JL. Using artificial intelligence methods to speed up drug discovery. Expert Opin Drug Discov 2019 Aug;14(8):769-777. [doi: 10.1080/17460441.2019.1621284] [Medline: 31140873]
9. Watson DS, Krutzinna J, Bruce IN, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ 2019 Mar 12;364:l886. [doi: 10.1136/bmj.l886] [Medline: 30862612]
10. Hamdi T, Ben Ali J, Di Costanzo V, Fnaiech F, Moreau E, Ginoux JM. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. Biocybern Biomed Eng 2018;38(2):362-372. [doi: 10.1016/j.bbe.2018.02.005]
11. Birjais R, Mourya AK, Chauhan R, Kaur H. Prediction and diagnosis of future diabetes risk: a machine learning approach. SN Appl Sci 2019 Sep;1(9):1-8. [doi: 10.1007/s42452-019-1117-9]
12. Sadhu A, Jadli A. Early-stage diabetes risk prediction: a comparative analysis of classification algorithms. IARJSET 2021;8(2):193-201. [doi: 10.17148/IARJSET.2021.8228]
13. Xue J, Min F, Ma F. Research on diabetes prediction method based on machine learning. In: In Journal of Physics: Conference Series: IOP Publishing; 2020, Vol. 1684:012062.
14. Le TM, Vo TM, Pham TN, Dao SVT. A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access 2020;9:7869-7884. [doi: 10.1109/ACCESS.2020.3047942] [Medline: 32542175]
15. Shafi MKI, Sultan MR, Rahman SMM, Hoque MM. IoT based smart home: a machine learning approach. Presented at: 2021 24th International Conference on Computer and Information Technology (ICCIT); Dec 18-20, 2021; Dhaka, Bangladesh p. 1-6. [doi: 10.1109/ICCIT54785.2021.9689786]
16. Xie J, Wang Q. Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-Series models. IEEE Trans Biomed Eng 2020 Nov;67(11):3101-3124. [doi: 10.1109/TBME.2020.2975959] [Medline: 32091990]
17. Kriukova G, Shvai N, Pereverzyev SV. Application of regularized ranking and collaborative filtering in predictive alarm algorithm for nocturnal hypoglycemia prevention. Presented at: 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems; Sep 21-23, 2017; Bucharest, Romania p. 634-638. [doi: 10.1109/IDAACS.2017.8095169]
18. Mujahid O, Contreras I, Vehi J. Machine learning techniques for hypoglycemia prediction: trends and challenges. Sensors (Basel) 2021 Jan 14;21(2):546. [doi: 10.3390/s21020546] [Medline: 33466659]
19. Vu L, Kefayati S, Idé T, et al. Predicting nocturnal hypoglycemia from continuous glucose monitoring data with extended prediction horizon. AMIA Annu Symp Proc 2020;2019:874-882. [Medline: 32308884]
20. Fathi AE, Breton MD. Using reinforcement learning to simplify mealtime insulin dosing for people with type 1 diabetes: in-silico experiments. IFAC-PapersOnLine 2023;56(2):11539-11544. [doi: 10.1016/j.ifacol.2023.10.446]
21. Tejedor M, Hjerde SN, Myhre JN, Godtliebsen F. Evaluating deep Q-learning algorithms for controlling blood glucose in in silico type 1 diabetes. Diagnostics (Basel) 2023 Oct 7;13(19):3150. [doi: 10.3390/diagnostics13193150] [Medline: 37835893]
22. Magni L, Raimondo DM, Bossi L, et al. Model predictive control of type 1 diabetes: an in silico trial. J Diabetes Sci Technol 2007 Nov;1(6):804-812. [doi: 10.1177/193229680700100603] [Medline: 19885152]
23. Yu X, Yang Z, Sun X, et al. Deep reinforcement learning for automated insulin delivery systems: algorithms, applications, and prospects. AI 2025;6(5):87. [doi: 10.3390/ai6050087]
24. Gottesman O, Johansson F, Komorowski M, et al. Guidelines for reinforcement learning in healthcare. Nat Med 2019 Jan;25(1):16-18. [doi: 10.1038/s41591-018-0310-5] [Medline: 30617332]

XSL•FO
RenderX

25. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017. [doi: 10.48550/arXiv.1705.07874]
26. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": explaining the predictions of any classifier. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA p. 1135-1144. [doi: 10.1145/2939672.2939778]
27. Al-Hamadani MNA, Fadhel MA, Alzubaidi L, Balazs H. Reinforcement learning algorithms and applications in healthcare and robotics: a comprehensive and systematic review. Sensors (Basel) 2024 Apr 11;24(8):2461. [doi: 10.3390/s24082461] [Medline: 38676080]
28. Marling C, Bunescu R. The OhioT1DM dataset for blood glucose level prediction: update 2020. CEUR Workshop Proc 2020 Sep;2675:71-74. [Medline: 33584164]
29. Marling C, Bunescu RC. The OhioT1DM dataset for blood glucose level prediction: update 2020. CEUR Workshop Proc 2020 Sep;2675:71-74. [Medline: 33584164]
30. Sutton RS, Barto AG. Reinforcement Learning: An Introduction, 2nd edition: MIT Press; 2018.
31. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. Int J Rob Res 2013 Sep;32(11):1238-1274. [doi: 10.1177/0278364913495721]
32. Berny A. Selection and reinforcement learning for combinatorial optimization. Presented at: Parallel Problem Solving from Nature PPSN VI: 6th International Conference; Sep 18-20, 2000; Paris, France p. 601-610. [doi: 10.1007/3-540-45356-3_59]
33. van Eck NJ, van Wezel M. Application of reinforcement learning to the game of Othello. Comput Oper Res 2008 Jun;35(6):1999-2017. [doi: 10.1016/j.cor.2006.10.004]
34. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
35. Oroojeni Mohammad Javad M, Agboola SO, Jethwani K, Zeid A, Kamarthi S. A reinforcement learning-based method for management of type 1 diabetes: exploratory study. JMIR Diabetes 2019 Aug 28;4(3):e12905. [doi: 10.2196/12905] [Medline: 31464196]
36. Sammut C, Webb GI. Encyclopedia of Machine Learning: Springer; 2011. [doi: 10.1007/978-0-387-30164-8]
37. Salkind NJ. Encyclopedia of Research Design: SAGE Publications, Inc; 2010.
38. Aleppo G. Clinical application of time in range and other metrics. Diabetes Spectr 2021 May;34(2):109-118. [doi: 10.2337/ds20-0093] [Medline: 34149251]
39. OhioT1DM dataset. University of North Carolina at Charlotte. URL: https://webpages.charlotte.edu/rbunescu/data/ohiot1dm/OhioT1DM-dataset.html [accessed 2026-05-23]

Abbreviations

AI: artificial intelligence
CGM: continuous glucose monitoring
DQN: deep Q-network
IOB: insulin on-board
LIME: local interpretable model-agnostic explanations
LR: logistic regression
LSTM: long short-term memory
MAE: mean absolute error
RL: reinforcement learning
RMSE: root mean square error
SHAP: Shapley additive explanations
T1D: type-1 diabetes
TIR: time in range

Edited by S Li; submitted 16.Jun.2025; peer-reviewed by K Braune; revised version received 29.Nov.2025; accepted 03.Dec.2025; published 03.Jun.2026.

Please cite as:
Taku EM, Gupta V, Singhal A
Personalized Type 1 Diabetes Management: Reinforcement Learning–Based Insulin Dosing and Glucose Forecasting
JMIR Diabetes 2026;11:e79195
URL: https://diabetes.jmir.org/2026/1/e79195
doi:10.2196/79195

XSL•FO
RenderX

© Ernest M Taku, Vibhuti Gupta, Ashutosh Singhal. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 3.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integration of Continuous Glucose Monitoring With HbA1c to Improve the Detection of Prediabetes in Asian Individuals: Model Development Study

Michelle H Lee[1,2*], PhD; Shihui Jin[3*], PhD; Eveline Febriana[4], MSc; Maybritte Lim[5], BSc; Sonia Baig[1], PhD; Shahmir H Ali[3], PhD; Ian Yi Han Ang[3], PhD; Tze Ping Loh[6], MBBChBAO; Ashna Nastar[7], MBBChir; Kee Seng Chia[3], MD; Alice Pik-Shan Kong[8], MBChB, MD; Faidon Magkos[9], PhD; Alex R Cook[3], PhD; Sue-Anne Toh[1,2,10], MBBChir, MA, MSc

[1]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, 10 Medical Drive, Singapore, Singapore

[10]Department of Medicine, National University Hospital, National University Health System, Singapore, Singapore

[2]NOVI Health, Singapore, Singapore

[3]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[4]Department of Clinical and Translational Research Office, National Heart Centre, Singapore, Singapore

[5]Department of Radiation Oncology, National University Cancer Institute, Singapore, Singapore

[6]Department of Laboratory Medicine, National University Hospital, Singapore, Singapore

[7]Department of Medicine, Alexandra Hospital, Singapore, Singapore

[8]Department of Medicine and Therapeutics, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[9]Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark

[*]these authors contributed equally

Corresponding Author:
Sue-Anne Toh, MBBChir, MA, MSc
Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, 10 Medical Drive, Singapore, Singapore

Abstract

Background: Glycated hemoglobin (HbA_{1c}) is a convenient tool to evaluate glycemic status but its ability to detect individuals at risk for type 2 diabetes is limited.

Objective: Exploiting the glycemic variability captured in continuous glucose monitoring (CGM), we used a well-characterized Asian cohort study from Singapore to assess whether utilizing CGM features in a machine learning model can improve the detection of prediabetes as compared to using HbA_{1c} alone.

Methods: In this study, 406 nondiabetic Asian participants underwent an oral glucose tolerance test and had their fasting and 2-hour plasma glucose concentrations measured, together with HbA_{1c}, to classify them as with normoglycemia or prediabetes. They also wore a CGM sensor for 14 days. CGM profile features were extracted and prediction models were constructed with random subsampling validation to evaluate predictive efficacy. The use of CGM and HbA_{1c} data alone or in combination was assessed for the ability to correctly distinguish prediabetes from normoglycemia.

Results: In this cohort (N=406), 189 (46.6%) individuals had prediabetes. The majority of the cohort were women (n=236, 58.1%) and of Chinese ethnicity (n=267, 65.8%). Those with prediabetes were slightly older, heavier, and had higher glucose levels with more variability than the normoglycemia group. A 2-step approach was used where those with HbA_{1c} ≥5.7% were automatically categorized as having prediabetes; the model then focused on the prediction capability of the CGM features among individuals with HbA_{1c} <5.7%. The prediction models with CGM outperformed the benchmark for comparison defined by HbA_{1c} ≥5.7%, where they yielded an area under the receiver operating characteristic curve of 0.866 - 0.876, with a lower specificity of 78% - 80% but a vastly improved sensitivity of 76% - 78%.

Conclusions: Adding CGM to HbA_{1c} in a 2-step approach greatly improved the sensitivity of detecting prediabetes in an Asian population. Given the benefits to optimizing lifestyle behaviors and its growing acceptability among the nondiabetic population, CGM is a promising alternative for type 2 diabetes mellitus risk screening.

Trial Registration: ClinicalTrials.gov NCT02838693; https://clinicaltrials.gov/NCT02838693

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e81520) doi:10.2196/81520

KEYWORDS

continuous glucose monitoring; HbA1c; machine learning; screening; Asian; prediabetes; glycated hemoglobin

Introduction

Glycated hemoglobin (HbA_{1c}) is the gold standard test for blood glucose monitoring and the management of diabetes in clinical practice, after 2 landmark prospective clinical studies (US Diabetes Control and Complications Trial [1] and the UK Prospective Diabetes Study [2]) clearly demonstrated a link between HbA_{1c} values and diabetes-related long-term complications, underpinning the use of HbA_{1c} as a surrogate marker for diabetes outcomes [3]. In 2010, HbA_{1c} was incorporated into the American Diabetes Association (ADA) criteria as a screening tool for prediabetes [4]. The advantages of HbA_{1c} include the absence of the requirement for fasting or timed blood samples, and its relative stability compared with glucose concentrations of fasting and 2 hours after an oral glucose tolerance test (OGTT), thereby reflecting long-term (~3 months) glycemic status. However, HbA_{1c} has a limited sensitivity to detect individuals at high risk of type 2 diabetes mellitus (T2DM) [5]. Data from 5395 nondiabetic individuals from the National Health and Nutrition Examination Survey showed that the current 5.7% cut-off has low sensitivity in detecting prediabetes, implying that HbA_{1c} values below 5.7% do not reliably exclude the presence of prediabetes [6]. Moreover, HbA_{1c} can be affected by blood disorders such as anemia [7], but also ethnicity [8], and does not provide any information on glycemic variability (ie, fluctuations in glucose levels over the course of a day).

The continuous glucose monitoring (CGM) system measures glucose levels subcutaneously in the interstitial fluid at regular time intervals for about 1 - 2 weeks. Since the first version by Medtronic in 1999, CGM has revolutionized glucose management in diabetes where excursions in glucose can be detected almost instantaneously, providing real-time information to better optimize medications for glucose control [9]. With the improvements in CGM technology, ease of use, and possible integration with other wearable biosensors, CGM is no longer confined to diabetes management. It is attracting the attention of healthy, nondiabetic individuals who want to assess their risk for T2DM or optimize their health [10]. Advances in data science and the application of machine learning techniques [11] have made it possible to exploit the glycemic variability captured in the CGM data to identify distinct patterns of glucose dysregulation for targeted clinical action [12,13], as well as distinguish those who are at risk of T2DM from those who are not [14-16].

Population-based screening can identify those who are at risk of T2DM (ie, prediabetes) to allow for early intervention and thereby mitigate the growing burden of the diabetes epidemic. This is critically important as almost one-in-two adults (45% or 240 million globally) living with diabetes (20 - 79 years old) are unaware of their status [17]. Recommended criteria for diagnosing prediabetes and T2DM encompass cut-offs for HbA_{1c}, but also fasting and 2-hour OGTT glucose concentrations [18], even though the OGTT is tedious and time-consuming, thereby reducing the effectiveness of a population-wide screening program. However, using HbA_{1c} alone is not sensitive enough to capture individuals at risk. There have been recent reports utilizing CGM data as a potential screening tool to discriminate between healthy individuals and those at risk of T2DM in European, Caucasian [14,16] and Indian [15] populations, but studies are scarce, especially in Asia, where 60% of all diabetes reside [19]. In this study, we took advantage of a well-characterized Asian cohort study from Singapore to assess whether utilizing CGM features in a machine learning model can improve the identification of prediabetes compared with HbA_{1c} alone.

Methods

Study Participants

This is a substudy anchored within the “Assessing the Progression to Type-2 Diabetes” (APT-2D) study (ClinicalTrials.gov: NCT02838693), which follows up on a large cohort of nondiabetic individuals for 3 years or until they develop T2DM [20].

For the main APT-2D study, healthy participants between the ages 30 and 70 years old, who were not on any long-term medication and had no prior history of diabetes, were recruited from April 2016 to December 2018. Recruitment was done through random sampling via various outreach efforts to the grassroots communities and organizations, and through media releases in order to recruit at least 2300 participants, so as to ensure adequate sample size for the cohort study outcomes.

For this substudy, participants with their penultimate visit (75 g OGTT) being conducted within 3 months from enrolling to this substudy (“Continuous Glucose Monitoring to Assess Glucose Dysregulation in Progression to Type-2 Diabetes” [CGM-APT2D]) and willing to wear a CGM sensor for 14 days were invited to participate. Out of the 449 individuals enrolled in this substudy, 429 participants contributed with CGM data. To narrow the study’s focus to the Asian population in this region, individuals from Europe or the Middle East were excluded from the analysis. Additionally, those diagnosed with T2DM during the study were removed to align with the study’s objective of examining individuals without diabetes. At the end, data from 419 individuals were used.

Ethical Considerations

The main study (APT-2D) was approved by the Domain Specific Review Board of the National Health Group (ref: 2016/00096) in Singapore. This substudy (CGM-APT2D) was approved by the Domain Specific Review Board of the National Healthcare Group (ref: 2020/01085) in Singapore. All participants provided written informed consent prior to joining the studies.

XSL•FO
RenderX

Experimental Procedures

OGTT

Participants were admitted after having fasted overnight for 10 - 12 hours. They were instructed to abstain from performing any strenuous exercise during the previous day and from consuming fat-rich foods during the preceding 3 days, to avoid potential delayed metabolic effects of exercise and high-fat feeding on metabolism. Participants arrived in the morning (8 AM); height, weight, and vital signs (heart rate and blood pressure) were obtained by standard methods after 5 minutes of rest, before any testing began. Participants then underwent a 2-hour OGTT. An indwelling catheter was inserted into an antecubital vein of one arm for blood sampling. A fasting blood sample was obtained at t=0 minutes, and then participants ingested a solution containing 75 g of glucose; additional blood samples were obtained at t=10, 20, 30, 60, 90 and 120 minutes for measurement of glucose concentrations. HbA_{1c} was measured using the fasting blood sample. Prediabetes and T2DM were defined by using the ADA criteria [18] for fasting and 2-hour plasma glucose and HbA_{1c} as previously described [20]. Participants classified as having newly diagnosed diabetes were excluded from the analysis.

CGM

A CGM sensor (Abbott Freestyle Libre) was attached to the upper arm to monitor interstitial glucose levels at 15-minute intervals for 14 days. The Freestyle Librelink app (Abbott) was used on participants' smartphones to scan and record their glucose readings. If a participant did not have a compatible smartphone, a reader was loaned to the participant to record glucose readings. Participants were instructed to scan the sensor at least once every 8 hours and to keep the sensor on for 14 days. Participants who had their sensors detached were offered a replacement sensor. The CGM sensor was worn by the participants within 3 months after completing the OGTT visit.

Sample Analysis

HbA_{1c} was measured in whole blood by cation-exchange high performance liquid chromatography (Bio-Rad Variant II Turbo; Bio-Rad). Plasma glucose during the OGTT was determined by the AU5822 general chemistry analyzer (Beckman Coulter) at the National University Hospital Referral Laboratory (accredited by the College of American Pathologists).

Data and Statistical Analyses

For each participant, the data encompassed 3 main components: demographic information, including age, sex, and ethnicity; clinical measurements recorded during the clinic visit, consisting of BMI, waist-hip ratio, fasting and 2-hour OGTT glucose (mmol/L), and HbA_{1c} (%); and glucose profile data for about 14 days obtained from the worn CGM sensor. The CGM data comprised a time series capturing the time and corresponding glucose levels (mmol/L), spaced at approximately 15-minute intervals. These intervals are not strictly equal, and the recording length and timing varied among individuals. Participants scanned the CGM sensor at an average frequency of 9.1 times a day. There were additional heterogeneities in the CGM datasets. For instance, occasional involuntary detachment of the sensor resulted in missing data, while the malfunctioning of sensors in some participants resulted in them wearing more than 1 sensor during the study period. Of the 429 participants, 8% required 1 replacement sensor and 0.5% required 2 replacement sensors.

For ease of comparison, we followed the work by Mao et al [13] and converted the CGM measurements into meaningful features that capture the essential characteristics of an individual's glucose trajectory. These features encompass measures of centrality (mean) and spread (maximum, minimum, standard deviation, coefficient of variation, and mean amplitude of glycemic excursion), along with proportions of time spent within abnormal (>7.8 or <3 mmol/L) and normal (3 to 7.8 mmol/L) glucose ranges. We additionally considered measures of glucose excursion, including the average rise and fall and their corresponding rates. Further details about the various CGM features, in terms of how they were defined and derived, are provided in Multimedia Appendix 1.

These summary statistics were computed after excluding the first 24 hours in each consecutive block of CGM recordings, as recommended by the manufacturer because glucose readings on the first day were typically lower compared to subsequent days. In cases where participants wore more than 1 device, we merged CGM recordings from different blocks after eliminating data from the "warm-up" period, given the negligible time gaps between them. During this process, 11 participants were subsequently removed due to insufficient CGM data captured by their devices, preventing the calculation of meaningful CGM summary statistics. Two participants were removed due to invalid HbA_{1c} values. Consequently, the final dataset comprises information from 406 participants with an average CGM recording length of 12 days.

We categorized these 406 participants based on their diabetic status into those with prediabetes and those with normoglycemia and prediabetes, thereby enabling a comparative analysis between groups. Specifically, for categorical variables such as sex and ethnicity, we used the Pearson chi-square test, while for continuous variables, we explored potential differences in group means through nonparametric Mann-Whitney U tests. P values were computed and a significance level of .05 was applied for interpretation.

Prediction models were then constructed to assess individual prediabetes risk. We explored three distinct sets of predictors in our analysis (Table S1 in Multimedia Appendix 1):

1. Demographic data encompassing age and gender, and 2 fundamental clinical measurements, namely BMI and waist-hip ratios (henceforth referred to as "Demo");
2. A combination of all the CGM summary statistics and the Demo data specified in set 1 ("CGM");
3. The CGM dataset listed in set 2, supplemented with the numerical HbA_{1c} value ("HbA_{1c}").

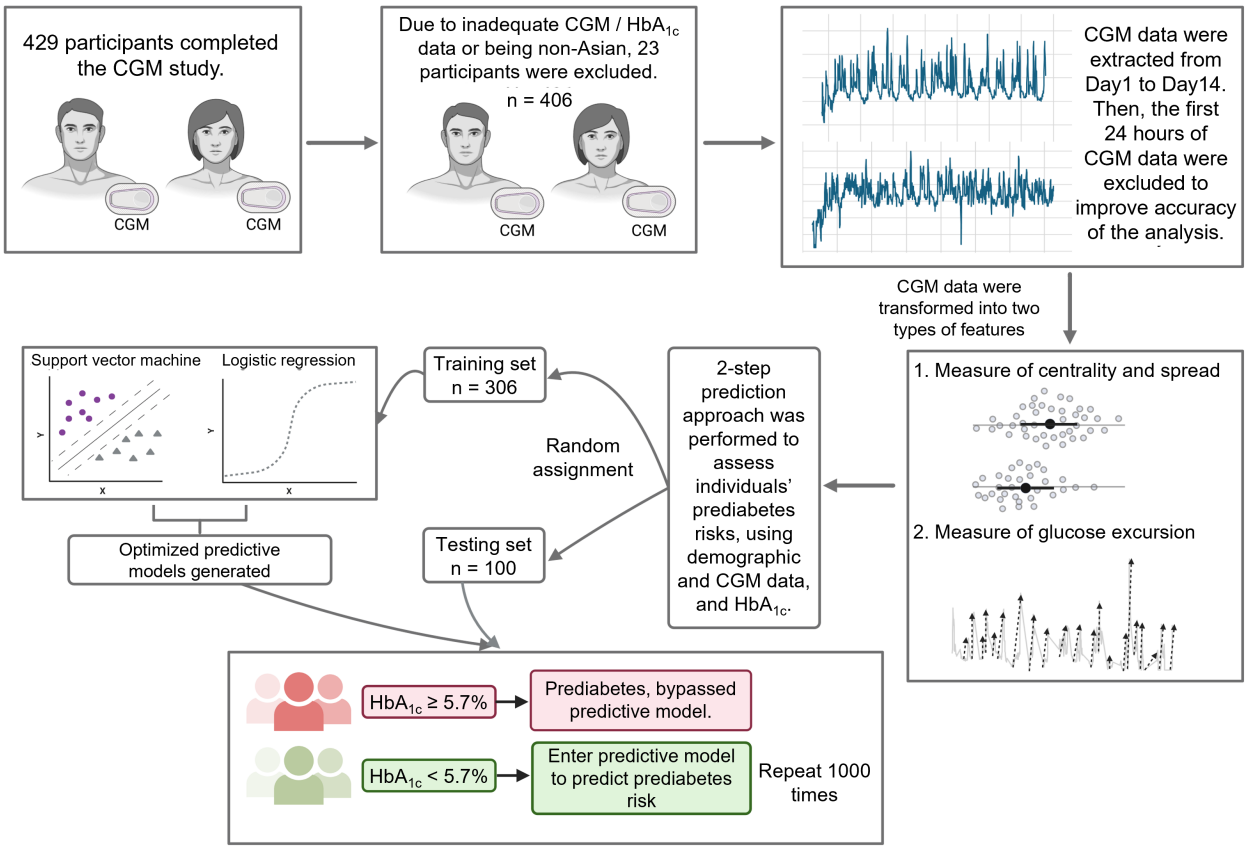
The performances of these models were compared with HbA_{1c} ≥5.7% alone to classify prediabetes as the benchmark for comparison.

XSL•FO
RenderX

For each of these data configurations, we employed 2 classification algorithms: logistic regression (LR) and support vector machine (SVM). To address potential concerns of overfitting, we implemented repeated random subsampling validation strategies, and evaluated the average prediction performance across 1000 randomly selected test sets, utilizing key metrics such as misclassification rates, specificity, and sensitivity.

In addition to our primary prediction models, we introduced a specialized 2-step prediction strategy. This approach categorized individuals with HbA_{1c} ≥5.7% as having prediabetes and focused exclusively on predicting the prediabetes risk among individuals with HbA_{1c} <5.7%. The predictor sets employed for this targeted analysis included the CGM dataset (set 2) and the HbA_{1c} dataset (set 3). We explored whether this segregation would enhance risk prediction process, allowing for a more nuanced understanding of prediabetes risk among individuals with relatively low HbA_{1c} levels. A sensitivity analysis was additionally conducted regarding the segregation threshold of 5.7%, but the alternative thresholds failed to yield better prediction accuracy in terms of misclassification rates (Tables S2–S3 in Multimedia Appendix 1). An overview of the methods is summarized in Figure 1.

Figure 1. Graphical overview of methods and analysis of CGM data. CGM: continuous glucose monitoring; HbA_{1c}: glycated hemoglobin.



Results

Participant Demographics and Baseline Clinical and CGM Measurements

Demographic information and clinical glucose measurements of the 406 Asian participants without diabetes are summarized in Table 1. Within this cohort, 189 (46.6%) individuals were categorized as prediabetic. The majority of the cohort were women (n=236, 58.1%) and individuals of Chinese ethnicity (n=267, 65.8%), with no significant differences observed in terms of sex and ethnic group distribution between the normoglycemia and prediabetes groups. The age of participants spanned from 33 to 74 years, with half falling between 42 and 57 years old. Those in the prediabetes group were ~6 years older and had a slightly greater BMI and waist-to-hip ratio than those in the normoglycemia group; the prevalence of obesity was 32.3% and 19.8% respectively, under the Singapore classification system [21]. Fasting and 2-hour OGTT glucose levels and HbA_{1c} were significantly greater in prediabetic than normoglycemic individuals (Table 1 and Figure S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Table . Demographics and baseline clinical glucose measurements in Asians without diabetes. Comparison between normoglycemia and prediabetes groups was performed using the Pearson chi-square test for categorical variables (sex and ethnicity), and the Mann-Whitney *U* test for continuous variables (ie, all but sex and ethnicity).

Variable	Total (N=406)	Normoglycemia (n=217)	Prediabetes (n=189)	*P* value
Sex, n (%)				.40
Male	170 (41.9)	95 (43.8)	75 (39.7)	
Female	236 (58.1)	122 (56.2)	114 (60.3)	
Ethnicity, n (%)				.46
Chinese	267 (65.8)	143 (65.9)	124 (65.6)	
Malay	57 (14.0)	27 (12.4)	30 (15.9)	
Indian	63 (15.5)	34 (15.7)	29 (15.3)	
Others	19 (4.7)	13 (6.0)	6 (3.2)	
Age (years), median (IQR)	49 (42-57)	46 (40-55)	52 (46-60)	<.001
BMI (kg/m^2), median (IQR)	24.7 (21.9-27.6)	24.0 (22.0-27.0)	25.0 (22.0-29.0)	.01
Waist-to-hip ratio, median (IQR)	0.87 (0.82-0.91)	0.86 (0.81-0.90)	0.88 (0.83-0.92)	.009
Fasting glucose (mmol/L), median (IQR)	5.0 (4.6-5.2)	4.8 (4.6, 5.1)	5.2 (4.8-5.4)	<.001
2-hour OGTT[a] glucose (mmol/L), median (IQR)	6.9 (5.8-8.2)	6.1 (5.2-6.9)	8.3 (7.2-9.3)	<.001
HbA_{1c}[b] (%), median (IQR)	5.5 (5.3-5.7)	5.3 (5.2-5.5)	5.7 (5.5-5.9)	<.001

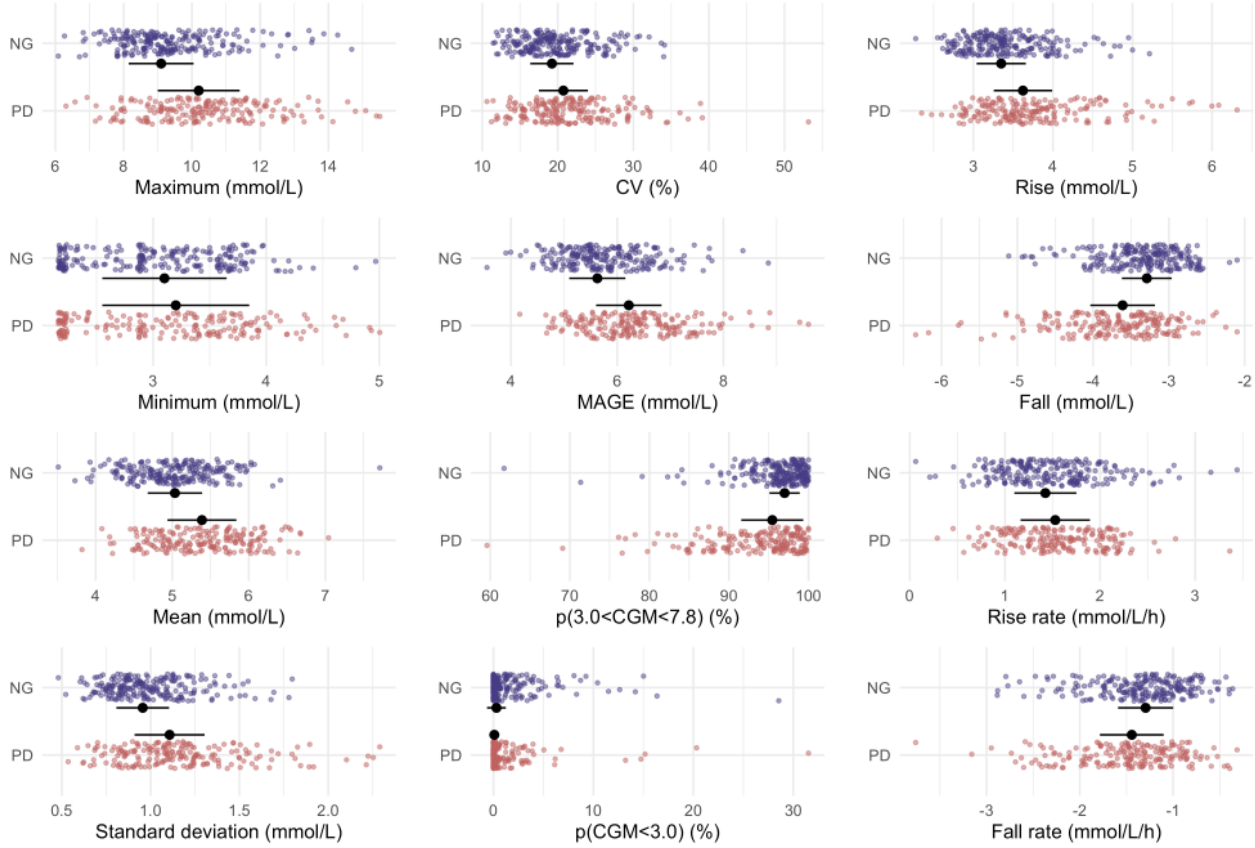
[a]OGTT: oral glucose tolerance test.

[b]HbA_{1c}: glycated hemoglobin.

With respect to glucose features derived from the CGM data, the prediabetes group displayed larger variability in glucose levels over time compared to the normoglycemia group, substantiated by significant differences in group means for most measurements related to spread and glucose excursion; only the minimum glucose was not significantly different between the 2 groups (Figure 2 and Table S4 in Multimedia Appendix 1). Notably, our assessment of time-in-range focused on the interval of 3.0 - 7.8 mmol/L. This choice stemmed from the consideration that the conventional cut-off for time-in-range (3.9 - 10 mmol/L) did not appear pertinent in this nondiabetic population, given the minimal divergence in the proportions of time spent outside this interval between the 2 groups (Table S5 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 2. Continuous glucose monitoring (CGM) glucose metrics in Asians with normoglycemia (NG) and prediabetes (PD). Significant differences were observed in all metrics except for the minimum glucose. CV: coefficient of variation; MAGE: mean amplitude of glycemic excursion.



Feasibility of Diagnosing Prediabetes Utilizing CGM Features

We set the benchmark for comparison using the prediabetes cut-off defined by HbA_{1c} ≥5.7%, resulting in a specificity of 100%, sensitivity of 60%, and a misclassification rate of 18%. The Demo model, which incorporated basic demographic information and obesity-related measures (age, gender, BMI, and waist-to-hip ratio) that are known risk factors for T2DM (Table S1 in Multimedia Appendix 1), demonstrated limited efficacy in distinguishing prediabetes from normoglycemia (Figure 3 and Table 2). The predictive performance greatly improved with the inclusion of CGM features in the model (CGM model), with a prediction sensitivity of 60% - 63% becoming comparable to that using HbA_{1c} ≥5.7% (Table 2). The diagnostic capability further increased when the HbA_{1c} value was added as a predictor to the CGM model (HbA_{1c} model), yielding a higher specificity of 78% - 80% and a sensitivity of 71% - 74%, and hence a substantially lower misclassification rate of 23% - 25% (Table 2). Notably, the sensitivity achieved by the HbA_{1c} prediction model surpassed that obtained when using HbA_{1c} ≥5.7% as a single threshold of 60%. In addition, the choice of classification algorithms did not fundamentally impact the predictive performance for all 3 models (Figure 3 and Table 2).

XSL•FO
RenderX

Figure 3. Receiver operating characteristic curves of the 3 model variants utilizing (A) logistic regression or (B) support vector machine as the classification algorithm. Model Demo: age, gender, BMI, and waist-hip ratio; Model CGM: Demo + CGM features; Model HbA_{1c}: Demo + CGM + HbA_{1c}. CGM: continuous glucose monitoring; HbA_{1c}: glycated hemoglobin.

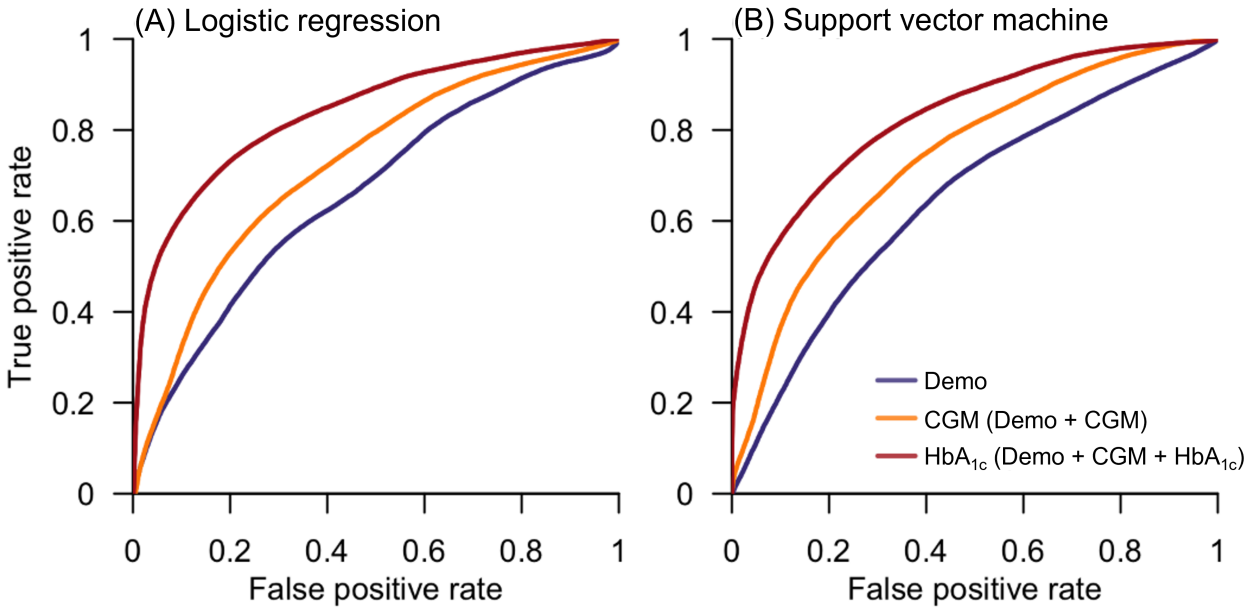


Table . Prediction accuracy of prediabetes, in terms of misclassification rates, specificity, and sensitivity, for model variants using distinct predictor sets and classification algorithms. The data are presented as the mean and 95% confidence intervals, derived from 1000 random splits into training and testing sets.

Model	Misclassification (%)	Specificity (%)	Sensitivity (%)	ROC AUC[b]
Demo, mean (95% CI)				
LR[c]	37.1 (36.8-37.4)	70.7 (70.4-71.1)	53.7 (53.3-54.2)	0.658 (0.654-0.661)
SVM[d]	38.0 (37.7-38.3)	71.6 (71.3-72.0)	50.9 (50.4-51.3)	0.648 (0.644-0.651)
CGM[e], mean (95% CI)				
LR	32.3 (32.0-32.6)	74.4 (74.1-74.8)	59.9 (59.5-60.3)	0.723 (0.719-0.726)
SVM	31.9 (31.6-32.2)	72.7 (72.3-73.1)	62.8 (62.4-63.3)	0.740 (0.736-0.743)
HbA_{1c}[f], mean (95% CI)				
LR	23.2 (23.0-23.5)	79.5 (79.1-79.8)	73.6 (73.2-74.0)	0.838 (0.836-0.841)
SVM	25.2 (24.9-25.4)	77.9 (77.5-78.2)	71.3 (70.9-71.7)	0.829 (0.826-0.831)
HbA_{1c} ≥5.7% cut-off (benchmark)	18.5	100	60.3	0.802

[a]Model Demo: age, gender, BMI and waist-hip ratio; Model CGM: Demo + CGM features; Model HbA_{1c}: Demo + CGM + HbA_{1c}.

[b]ROC AUC: area under the receiver operating characteristic curve.

[c]LR: logistic regression.

[d]SVM: support vector machine.

[e]CGM: continuous glucose monitoring.

[f]HbA_{1c}: glycated hemoglobin.

Enhancing Prediction Efficacy Through the 2-Step Approach

We extended our analysis to assess whether stratifying the population based on their HbA_{1c} levels could improve the detection of prediabetes using a 2-step approach where those with HbA_{1c} ≥5.7% were automatically categorized as having prediabetes, thereby focusing the prediction capability on those with HbA_{1c} <5.7%. Overall, this 2-step approach outperformed the benchmark for comparison using the prediabetes cut-off defined by HbA_{1c} ≥5.7%, and significantly improved the predictive performance in the HbA_{1c} model of both algorithms where $LR_{ROC\ AUC}$ increased from 0.838 to 0.866 and $SVM_{ROC\ AUC}$ increased from 0.829 to 0.876 (Tables 2 and 3 and Figure

XSL•FO
RenderX

4). In the LR HbA_{1c} model, there was an increase in sensitivity from 73.6% to 75.7%, resulting in a net reduction of the misclassification rate from 23.2% to 22.3%. A similar trend was also observed when SVM was used (Tables 2 and 3). Interestingly, the inclusion of the HbA_{1c} value as a variable in the 2-step approach did not improve the predictive performance in both the LR (area under the receiver operating characteristic curve [ROC AUC] of CGM: 0.872 vs ROC AUC of HbA_{1c}: 0.866) and SVM (ROC AUC of CGM: 0.881 vs ROC AUC of HbA_{1c}: 0.876) models (Table 3). Assessment of predictive performance using the area under the precision recall curve yielded a similar outcome (Table 3). Nonlinear models such as random forest and Extreme Gradient Boosting were also explored (Table S6 in Multimedia Appendix 1) but were not superior in predictive performance compared to the LR and SVM models in the 2-step approach (Table S7 in Multimedia Appendix 1 vs Table 3). Thus, LR and SVM were chosen for their ease of interpretability and clinical usability.

Table . Prediction accuracy of prediabetes, including misclassification rate, specificity, and sensitivity, for all data points within the testing sets. This evaluation was conducted for model variants that used the 2-step prediction strategy and drew inference from all data points in the training sets, irrespective of their glycated hemoglobin (HbA_{1c}) levels. The data are presented as the mean and 95% confidence intervals, derived from 1000 random splits into training and testing sets.

Model[a]	Misclassification (%)	Specificity (%)	Sensitivity (%)	ROC AUC[b]	PRC AUC[c]
CGM[d], mean (95% CI)					
LR[e]	22.5 (22.2-22.7)	74.4 (74.1-74.8)	81.1 (80.7-81.4)	0.872 (0.869-0.874)	0.893 (0.891-0.895)
SVM[f]	23.1 (22.8-23.4)	72.7 (72.3-73.1)	81.8 (81.4-82.2)	0.881 (0.879-0.883)	0.899 (0.897-0.901)
HbA_{1c}, mean (95% CI)					
LR	22.3 (22.0-22.5)	79.5 (79.1-79.8)	75.7 (75.3-76.1)	0.866 (0.863-0.868)	0.889 (0.887-0.891)
SVM	21.9 (21.6-22.1)	77.9 (77.5-78.2)	78.4 (78.0-78.8)	0.876 (0.874-0.879)	0.898 (0.896-0.900)

[a]Model Demo: age, gender, BMI, and waist-hip ratio; Model CGM: Demo + CGM features; Model HbA1c: Demo + CGM + HbA1c.

[b]area under the receiver operating characteristic curve.

[c]area under the precision recall curve.

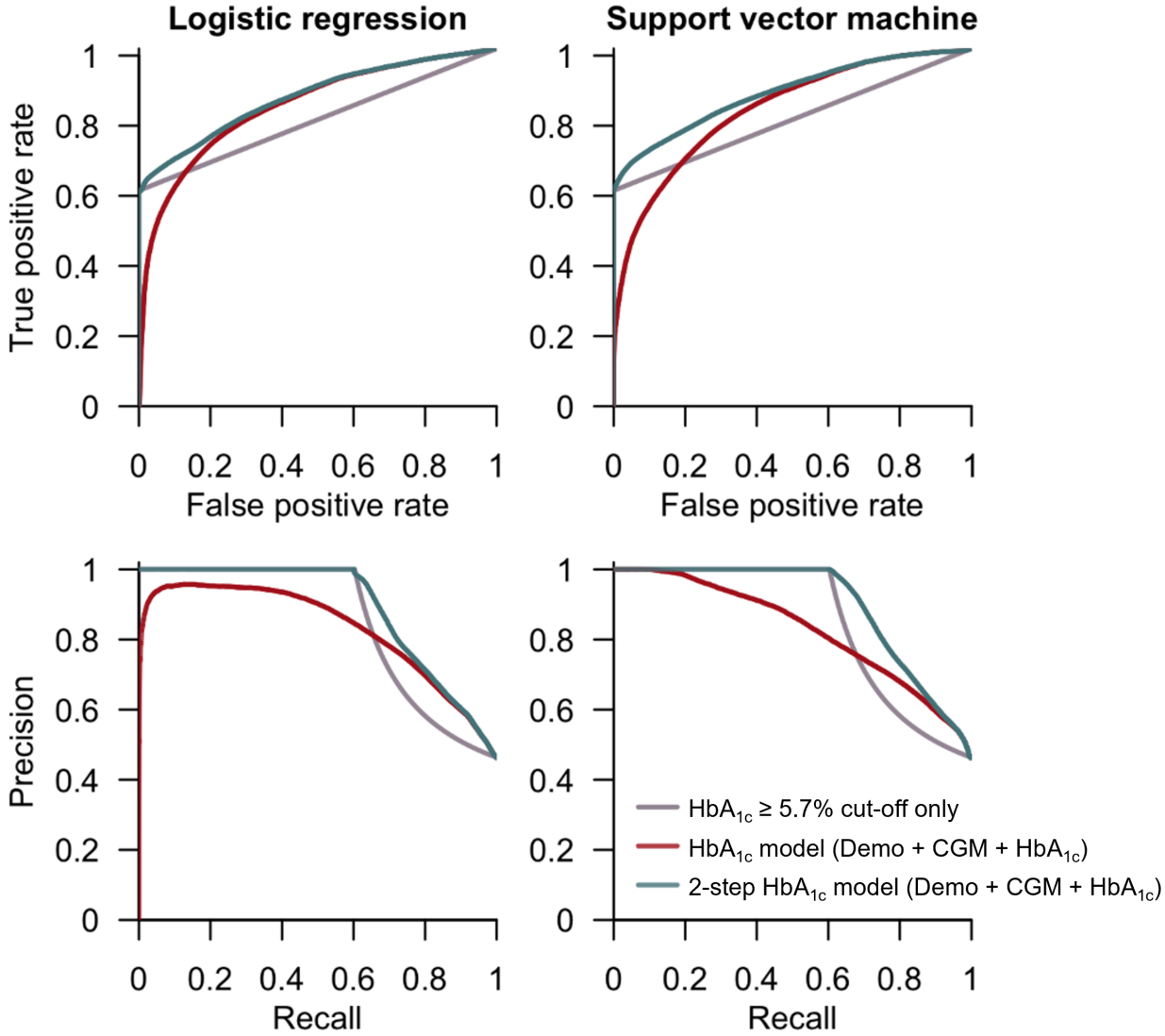
[d]CGM: continuous glucose monitoring.

[e]LR: logistic regression.

[f]SVM: support vector machine.

XSL•FO
RenderX

Figure 4. The receiver operating characteristic and precision recall curves comparing the original model HbA_{1c} with the 2-step approach utilizing logistic regression or support vector machine as the classification algorithm. HbA_{1c} ≥5.7% cut-off to classify prediabetes was used as the benchmark for comparison. CGM: continuous glucose monitoring; HbA_{1c}: glycated hemoglobin.



To further assess the feasibility of using CGM data to accurately identify individuals with prediabetes whose HbA_{1c} were in the normoglycemic range, we next focused on individuals with HbA_{1c} levels <5.7% (n=292) and compared the prediction outcomes for this subpopulation. We set the benchmark for comparison using the prediabetes cut-off defined by HbA_{1c} ≥5.7%, resulting in a specificity of 100%, sensitivity of 0%, and a misclassification rate of 25.7%. The ROC AUC and area under the precision recall curve values of the models were significantly higher than 0.5 and the test results yielded a slightly higher misclassification rate of 30% - 32%, a lower specificity of 73% - 80%, but a much higher sensitivity rate of 40% - 54% compared to the benchmark. SVM was slightly superior to LR in overall predictive performance and the omission of HbA_{1c} value as a variable resulted in a better model performance (Table 4).

XSL•FO
RenderX

Table . Prediction accuracy, including misclassification rates, specificity, and sensitivity, for those in the testing sets with glycated hemoglobin (HbA_{1c}) <5.7%. This evaluation was conducted for model variants that utilized the 2-step prediction strategy and drew inference from all data points in the training sets, irrespective of their HbA_{1c} levels. The data are presented as the mean and 95% confidence intervals, derived from 1000 random splits into training and testing sets.

Model[a]	Misclassification (%)	Specificity (%)	Sensitivity (%)	ROC AUC[b]	PRC AUC[c]
CGM[d], mean (95% CI)					
LR[e]	31.1 (30.8-31.5)	74.4 (74.1-74.8)	52.6 (51.9-53.3)	0.679 (0.674-0.683)	0.655 (0.650-0.659)
SVM[f]	32.0 (31.6-32.3)	72.7 (72.3-73.1)	54.4 (53.7-55.1)	0.703 (0.698-0.707)	0.640 (0.636-0.645)
HbA_{1c}, mean (95% CI)					
LR	30.8 (30.5-31.1)	79.5 (79.1-79.8)	39.2 (38.5-39.9)	0.663 (0.659-0.668)	0.657 (0.652-0.661)
SVM	30.3 (30.0-30.6)	77.9 (77.5-78.2)	45.9 (45.2-46.6)	0.690 (0.686-0.695)	0.634 (0.629-0.638)
HbA_{1c} ≥5.7% cut-off (benchmark)	25.7	100	0	0	0

[a]Model Demo: age, gender, BMI, and waist-hip ratio; Model CGM: Demo + CGM features; Model HbA_{1c}: Demo + CGM + HbA_{1c}.

[b]area under the receiver operating characteristic curve.

[c]area under the precision recall curve.

[d]CGM: continuous glucose monitoring.

[e]LR: logistic regression.

[f]SVM: support vector machine.

Discussion

Our findings demonstrate that combining CGM data together with HbA_{1c} greatly improved the sensitivity of detecting prediabetes. The best strategy was a 2-step approach where those with HbA_{1c} ≥5.7% were classified as having prediabetes, and the model was then used to detect those with prediabetes in the remaining population with HbA_{1c} <5.7% (who, otherwise, would have been classified as normoglycemic by HbA_{1c} alone). In this study, if HbA_{1c} alone was used to detect presence of prediabetes, of those who had HbA_{1c} <5.7%, 26% were prediabetic either by fasting and/or 2-hour plasma glucose, and would have been misclassified as normoglycemic. We feel that the compromise of losing specificity for a much higher sensitivity is a good strategy for prediabetes screening, as it will increase the chances of detecting those at risk of T2DM, while the false positives will still benefit from a diabetes prevention program that improves their lifestyle and diet choices. These findings can be implemented in a clinical setting, where general practitioners are able to improve HbA_{1c}'s detection of prediabetes by adding the CGM into their arsenal of screening tools.

Our model to identify prediabetes had a sensitivity of 81.8% and specificity of 72.7%, a performance comparable to other CGM modeling studies. Acciaroli et al [14] reported an 86% sensitivity in identifying those with impaired glucose tolerance in a Caucasian population, while Kaufman et al [15] reported an 86% sensitivity and a specificity of 71% - 78% in identifying those with prediabetes from a study in India. Comparing the 2 classification algorithms employed in our study, SVM performed slightly better than LR, and the addition of HbA_{1c} as a variable greatly improved the CGM model. This was as expected since HbA_{1c} is, by definition, one of the parameters for prediabetes classification in the ADA criteria used in this study [18]. However, in the 2-step approach where those with HbA_{1c} ≥5.7% were automatically categorized as having prediabetes, further addition of HbA_{1c} as a variable was no longer beneficial to the CGM model. While we performed robust internal validation through repeated random subsampling to minimize overfitting, we acknowledge this does not substitute for external validation in independent cohorts. External validation would be needed to assess generalizability to other Asian populations and health care settings.

While the current consensus for time-in-range among patients with diabetes is 3.9 - 10.0 mmol/L, there is an emerging secondary measure, termed time-in-tight-range of 3.9 - 7.8 mmol/L, which is believed to better represent normoglycemia [22,23]. Nondiabetic participants of Western descent have an average time-in-tight-range of 96% - 97% [24,25]. However, in a Chinese population, the corresponding proportion was only 93% [26]. Moreover, it has been reported that nondiabetic individuals have a nonnegligible time spent in the hypoglycemic range of <3.9 mmol/L, calling to question whether the 3.9 mmol/L cut-off is relevant in persons without diabetes [24,26,27]. In our population, the amount of time spent in <3.9 mmol/L was 2.9% with the range almost exclusively in 3.0 - 3.9 mmol/L. While we acknowledge that there have been reports of the Freestyle Libre CGM sensor underperforming in the hypoglycemic range [28], reducing the lower bound cut-off from 3.9 to 3.0 mmol/L supports the current recommendation of using 3.0 mmol/L to define clinically important hypoglycemia instead [22]. Hence, our recommended target time-in-range was 3.0 - 7.8 mmol/L, which captured an average of 96.3% of time spent in this glucose range in an Asian nondiabetic population.

As the CGM provides real-time biofeedback to the user, monitoring with the CGM would serve a dual purpose of

XSL•FO
RenderX

diagnosis and motivation for individuals to engage in healthy lifestyle behaviors, which in turn could improve glucose control. For prediabetic individuals, the use of CGM would be highly useful, as lifestyle modifications may be even more effective in reducing T2DM risk than metformin therapy (Diabetes Prevention Program) [29]. A short 10-day CGM along with exposure to activity and dietary insights, even without any specific dietary recommendations, was sufficient to significantly improve the time-in-range (3.0 - 7.8 mmol/L) in nondiabetic individuals [30]. However, we acknowledge that we did not assess behavioral changes of the participants in this study and this claim would require prospective validation. In recent times, there is an increasing acceptability of CGM among the public even in the absence of diabetes [31], rating it useful to improving their lifestyle [27]. Furthermore, CGM may represent a more acceptable alternative for diabetes screening than the current gold-standard OGTT. In a gestational diabetes screening study, the pregnant participants perceived CGM as significantly more acceptable than the OGTT [32].

The main strengths of our study are the substantial sample size (for a clinical study) in which the major Asian groups of East, South, and Southeast Asia were represented, and the OGTT in addition with HbA_{1c} were performed to accurately determine the participant's glycemic status. However, because the numbers of participants from other ethnicities such as Malay and Indian were low, we could not perform a subpopulation analysis to investigate whether the accuracy of the algorithm was affected by ethnicity. One limitation from our study design is that participants were not blinded to their own CGM data, which may have promoted healthier behaviors and improved their glucose profiles over the course of the recording period. Thus, the CGM profile may less accurately reflect the glycemic status from the OGTT and HbA_{1c} tests performed earlier (within 3 months). Our study population also did not include those with more severe hyperglycemia, that is, T2DM, hence it cannot provide a comprehensive discrimination between the varying degrees of glucose dysregulation and the possibility of detecting T2DM at the early stage of the disease. Lastly, while our study reported a prevalence of prediabetes at 47%, this is only 3% - 9% higher than other Asian studies with a Chinese-majority population, ranging from 38% - 44% [33,34]. For Yue et al [34] in particular, while their population demographics was approximately 10 years older than our study, their criteria for prediabetes only included impaired fasting glucose and HbA_{1c} 5.7% - 6.4%, while the impaired glucose tolerance was not taken into consideration; hence, the true prevalence for that study would have been higher than 44%. A single CGM assessment of 12 to 14 days in this cross-sectional study was sufficient to obtain adequate representation of the glucose variability that was distinguishable between the normoglycemia and prediabetes groups in this cohort. However, as CGM patterns may vary with seasonal changes, dietary patterns, or life circumstances, which a single assessment may not capture, future longitudinal work could examine whether repeated CGM assessments improve risk stratification or predict progression from prediabetes to T2DM.

In conclusion, the addition of CGM to HbA_{1c} in a 2-step approach, using either LR or SVM, greatly improved the sensitivity of detecting those at risk of T2DM in an Asian population. Given the benefits of the CGM to optimize lifestyle behaviors and its growing interest and acceptability among the nondiabetic population, CGM is an increasingly promising alternative to the classic OGTT for screening individuals at risk for T2DM in clinical practice.

Acknowledgments

The authors express their heartfelt thanks to the following: study participants for their participation, the Assessing the Progression to Type-2 Diabetes (APT-2D) study team for participant recruitment and conducting the clinical procedures, the Investigational Medicine Unit and Outpatient Parenteral Antibiotic Therapy unit of the National University Hospital for the usage of the clinical facilities, Saw Swee Hock School of Public Health of the National University of Singapore, and the National University Health System Regional Health System for their assistance in participant recruitment.

Funding

This work was supported by the National Medical Research Council (NMRC), Ministry of Health (MOH) Industry Alignment Fund [NMRC/MOHIAFCat1/0048/2016] and Janssen Pharmaceuticals Inc. (USA). In-kind contributions were supported by the National University Hospital (NUH) and the National University of Singapore (NUS). The funders had no role in the writing of the manuscript or the decision to submit it for publication. The funders had no role in the data collection, analysis or interpretation, 24 trial design or participant recruitment. The authors were not paid to write this article by a pharmaceutical company or any other agency. All authors had full access to the full data in the study and accept responsibility to submit for publication.

Data Availability

The data are not publicly available due to ethical restrictions but can be made available upon reasonable request with the submission of an appropriate research plan, and pending approval by the corresponding and senior authors.

Authors' Contributions

MHL, ML and SAT were involved in the study design, MHL, ML, EF and SAT were involved in the conduct of the study and data collection. SJ and ARC performed the prediction modellingmodeling, MHL, SJ and EF drafted the manuscript. All authors

XSL•FO
RenderX

were responsible for the data interpretation, critical review and intellectual input, and approving the final version of the manuscript. MHL and SJ are the guarantors of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Michelle H. Lee and Shihui Jin are co-first authors and contributed equally to the manuscript.

Conflicts of Interest

SAT is a board member and shareholder in NOVI Health (Singapore); has served on the advisory board for Novo Nordisk, Eli Lilly, Janssen, Boehringer-Ingelheim, Merck, Abbott and Astra Zeneca; is a member of the speaker's bureau for Eli Lilly, Boehringer-Ingelheim, DKSH, Merck, Abbott and Astra Zeneca; has received research support from Janssen, Merck and National Medical Research Council (NMRC), Singapore. MHL is an employee in NOVI Health (Singapore). APSK has received research grants and/or speaker honoraria from Abbott, Astra Zeneca, Bayer, Boehringer Ingelheim, Dexcom, Eli-Lilly, Kyowa Kirin, Merck Serono, Nestle, Novo-Nordisk, Pfizer and Sanofi. The remaining authors have no conflicts of interest to declare that are relevant to the content of this article. We recognize that financial relationships with continuous glucose monitor (CGM) manufacturers could theoretically influence study design or interpretation. To address this concern, we used gold-standard American Diabetes Association (ADA) criteria as our reference standard, followed pre-specified analytical protocols, and included co-authors without industry relationships in all critical analytical and interpretive decisions.

Multimedia Appendix 1
Supplementary methods and results.
[DOCX File, 385 KB - diabetes_v11i1e81520_app1.docx]

References

1. Diabetes Control and Complications Trial Research Group, Nathan DM, Genuth S, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
2. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):837-853. [doi: 10.1016/S0140-6736(98)07019-6] [Medline: 9742976]
3. Gillery P. A history of HbA1c through Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2013 Jan;51(1):65-74. [doi: 10.1515/cclm-2012-0548] [Medline: 22992284]
4. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010 Jan;33 Suppl 1(Suppl 1):S62-S69. [doi: 10.2337/dc10-S062] [Medline: 20042775]
5. Kodama S, Horikawa C, Fujihara K, et al. Use of high-normal levels of haemoglobin A(1C) and fasting plasma glucose for diabetes screening and for prediction: a meta-analysis. Diabetes Metab Res Rev 2013 Nov;29(8):680-692. [doi: 10.1002/dmrr.2445] [Medline: 23963843]
6. Guo F, Moellering DR, Garvey WT. Use of HbA1c for diagnoses of diabetes and prediabetes: comparison with diagnoses based on fasting and 2-hr glucose values and effects of gender, race, and age. Metab Syndr Relat Disord 2014 Jun;12(5):258-268. [doi: 10.1089/met.2013.0128] [Medline: 24512556]
7. Sinha N, Mishra TK, Singh T, Gupta N. Effect of iron deficiency anemia on hemoglobin A1c levels. Ann Lab Med 2012 Jan;32(1):17-22. [doi: 10.3343/alm.2012.32.1.17] [Medline: 22259774]
8. Bergenstal RM, Gal RL, Connor CG, et al. Racial differences in the relationship of glucose concentrations and hemoglobin A1c levels. Ann Intern Med 2017 Jul 18;167(2):95-102. [doi: 10.7326/M16-2596] [Medline: 28605777]
9. Olczuk D, Priefer R. A history of continuous glucose monitors (CGMs) in self-monitoring of diabetes mellitus. Diabetes Metab Syndr 2018;12(2):181-187. [doi: 10.1016/j.dsx.2017.09.005] [Medline: 28967612]
10. Holzer R, Bloch W, Brinkmann C. Continuous glucose monitoring in healthy adults-possible applications in health care, wellness, and sports. Sensors (Basel) 2022 Mar 5;22(5):2030. [doi: 10.3390/s22052030] [Medline: 35271177]
11. Lalani B, Herur R, Zade D, et al. Applications of artificial intelligence and machine learning in prediabetes: a scoping review. J Diabetes Sci Technol 2025 Jul 8;8:19322968251351995. [doi: 10.1177/19322968251351995] [Medline: 40626375]
12. Hall H, Perelman D, Breschi A, et al. Glucotypes reveal new patterns of glucose dysregulation. PLoS Biol 2018 Jul;16(7):e2005143. [doi: 10.1371/journal.pbio.2005143] [Medline: 30040822]
13. Mao Y, Tan KXQ, Seng A, Wong P, Toh SA, Cook AR. Stratification of patients with diabetes using continuous glucose monitoring profiles and machine learning. Health Data Sci 2022;2022:9892340. [doi: 10.34133/2022/9892340] [Medline: 38487483]
14. Acciaroli G, Sparacino G, Hakaste L, et al. Diabetes and prediabetes classification using glycemic variability indices from continuous glucose monitoring data. J Diabetes Sci Technol 2018 Jan;12(1):105-113. [doi: 10.1177/1932296817710478] [Medline: 28569077]
15. Kaufman JM, van Veen L, Fossat Y. Screening for impaired glucose homeostasis: a novel metric of glycemic control. Mayo Clin Proc Digit Health 2023 Jun;1(2):189-200. [doi: 10.1016/j.mcpdig.2023.02.008] [Medline: 40206722]

XSL•FO
RenderX

16. Cichosz SL, Kronborg T, Laugesen E, et al. From stability to variability: classification of healthy individuals, prediabetes, and type 2 diabetes using glycemic variability indices from continuous glucose monitoring data. Diabetes Technol Ther 2025 Jan;27(1):34-44. [doi: 10.1089/dia.2024.0226] [Medline: 39115921]
17. Diabetes Atlas, 10th edition: International Diabetes Federation; 2021.
18. American Diabetes Association Professional Practice Committee. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022. Diabetes Care 2022 Jan 1;45(Suppl 1):S17-S38. [doi: 10.2337/dc22-S002] [Medline: 34964875]
19. Nanditha A, Ma RCW, Ramachandran A, et al. Diabetes in Asia and the Pacific: implications for the global epidemic. Diabetes Care 2016 Mar;39(3):472-485. [doi: 10.2337/dc15-1536] [Medline: 26908931]
20. Magkos F, Lee MH, Lim M, et al. Dynamic assessment of insulin secretion and insulin resistance in Asians with prediabetes. Metab Clin Exp 2022 Mar;128:154957. [doi: 10.1016/j.metabol.2021.154957] [Medline: 34942192]
21. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
22. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023 Jan;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9] [Medline: 36493795]
23. Passanisi S, Piona C, Salzano G, et al. Aiming for the best glycemic control beyond time in range: time in tight range as a new continuous glucose monitoring metric in children and adolescents with type 1 diabetes using different treatment modalities. Diabetes Technol Ther 2024 Mar;26(3):161-166. [doi: 10.1089/dia.2023.0373] [Medline: 37902743]
24. Shah VN, DuBose SN, Li Z, et al. Continuous glucose monitoring profiles in healthy nondiabetic participants: a multicenter prospective study. J Clin Endocrinol Metab 2019 Oct 1;104(10):4356-4364. [doi: 10.1210/jc.2018-02763] [Medline: 31127824]
25. Rodriguez-Segade S, Rodriguez J, Camiña F, et al. Continuous glucose monitoring is more sensitive than HbA1c and fasting glucose in detecting dysglycaemia in a Spanish population without diabetes. Diabetes Res Clin Pract 2018 Aug;142:100-109. [doi: 10.1016/j.diabres.2018.05.026] [Medline: 29807103]
26. Zhou J, Li H, Ran X, et al. Reference values for continuous glucose monitoring in Chinese subjects. Diabetes Care 2009 Jul;32(7):1188-1193. [doi: 10.2337/dc09-0076] [Medline: 19389816]
27. Sofizadeh S, Pehrsson A, Ólafsdóttir AF, Lind M. Evaluation of reference metrics for continuous glucose monitoring in persons without diabetes and prediabetes. J Diabetes Sci Technol 2022 Mar;16(2):373-382. [doi: 10.1177/1932296820965599] [Medline: 33100059]
28. Zhou Y, Mai X, Deng H, et al. Discrepancies in glycemic metrics derived from different continuous glucose monitoring systems in adult patients with type 1 diabetes mellitus. J Diabetes 2022 Jul;14(7):476-484. [doi: 10.1111/1753-0407.13296] [Medline: 35864804]
29. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403. [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
30. Dehghani Zahedani A, Shariat Torbaghan S, Rahili S, et al. Improvement in glucose regulation using a digital tracker and continuous glucose monitoring in healthy adults and those with type 2 diabetes. Diabetes Ther 2021 Jul;12(7):1871-1886. [doi: 10.1007/s13300-021-01081-3] [Medline: 34047962]
31. Liao Y, Schembre S. Acceptability of continuous glucose monitoring in free-living healthy individuals: implications for the use of wearable biosensors in diet and physical activity research. JMIR Mhealth Uhealth 2018 Oct 24;6(10):e11181. [doi: 10.2196/11181] [Medline: 30355561]
32. Di Filippo D, Henry A, Bell C, et al. A new continuous glucose monitor for the diagnosis of gestational diabetes mellitus: a pilot study. BMC Pregnancy Childbirth 2023 Mar 18;23(1):186. [doi: 10.1186/s12884-023-05496-7] [Medline: 36932353]
33. Wang L, Peng W, Zhao Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021 Dec 28;326(24):2498-2506. [doi: 10.1001/jama.2021.22208] [Medline: 34962526]
34. Yue L, Tian Y, Ma M, et al. Prevalence of prediabetes and risk of CVD mortality in individuals with prediabetes alone or plus hypertension in Northeast China: insight from a population based cohort study. BMC Public Health 2024 Feb 15;24(1):475. [doi: 10.1186/s12889-024-17996-y] [Medline: 38360567]

Abbreviations

ADA: American Diabetes Association
APT-2D: Assessing the Progression to Type-2 Diabetes
CGM: continuous glucose monitoring
CGM-APT2D: Continuous Glucose Monitoring to Assess Glucose Dysregulation in Progression to Type-2 Diabetes
HbA_{1c}: glycated hemoglobin
LR: logistic regression
OGTT: oral glucose tolerance test

XSL•FO
RenderX

ROC AUC: area under the receiver operating characteristic curve
SVM: support vector machine
T2DM: type 2 diabetes mellitus

Edited by I Steenstra; submitted 30.Jul.2025; peer-reviewed by G Lim, M Amini; revised version received 09.Mar.2026; accepted 11.Mar.2026; published 27.Apr.2026.

Please cite as:
Lee MH, Jin S, Febriana E, Lim M, Baig S, Ali SH, Ang IYH, Loh TP, Nastar A, Chia KS, Kong APS, Magkos F, Cook AR, Toh SA Integration of Continuous Glucose Monitoring With HbA1c to Improve the Detection of Prediabetes in Asian Individuals: Model Development Study
JMIR Diabetes 2026;11:e81520
URL: https://diabetes.jmir.org/2026/1/e81520
doi:10.2196/81520

© Michelle H Lee, Shihui Jin, Eveline Febriana, Maybritte Lim, Sonia Baig, Shahmir H Ali, Ian Yi Han Ang, Tze Ping Loh, Ashna Nastar, Kee Seng Chia, Alice Pik-Shan Kong, Faidon Magkos, Alex R Cook, Sue-Anne Toh. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 27.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Continuous Ketone Monitoring: Data From a Randomized Controlled Trial

Simon K Kjær[1,2], MD; Lukas Ochsner Reynaud Ridder[2,3], MD, PhD; Mads Svart[2,4], MD, PhD; Nikolaj Rittig[1,2], MD, PhD; Lise Nørkjær Bjerg[5], MSc, PhD; Birgitte Sandfeld-Paulsen[6], MD, PhD; Henrik Holm Thomsen[2,7], MD, PhD

[1]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark
[2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[3]Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark
[4]Department of Internal Medicine, Regional Hospital Horsens, Horsens, Denmark
[5]Department of Clinical Biochemistry, Regional Hospital Central Jutland, Viborg, Denmark
[6]Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark
[7]Medical Diagnostic Center, University Clinic for Innovative Patient Pathways, Regional Hospital Central Jutland, Heibergs Allé 4K, Viborg, Denmark

Corresponding Author:
Henrik Holm Thomsen, MD, PhD
Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Abstract

In our study, a commercially available continuous ketone monitoring device captured β-Hydroxybutyrate (BHB) dynamics during exogenous ketosis but revealed a gradual decline day-to-day BHB concentrations over 14 days in both ketone ester and placebo groups, likely reflecting sensor drift.

(JMIR Diabetes 2026;11:e85548) doi:10.2196/85548

KEYWORDS

continuous ketone monitoring; exogenous ketosis; beta-hydroxybutyrate; ketone sensor technology; ketone ester

Introduction

Continuous measurement of ketone bodies is of scientific and clinical interest, providing insights into type 1 and type 2 diabetes, ketogenic diets, intermittent fasting, and exogenous ketone precursor supplementation. Current finger-prick point-of-care testing (POCT) devices are invasive, intermittent, and fail to capture dynamic fluctuations [1]. Continuous ketone monitoring (CKM), a small device measuring interstitial ketone (β-hydroxybutyrate, BHB) levels, offers a potential solution [2]. CKM research, however, remains in its early stages, with only a single commercially available device at present (SiBio KS1, Hong Kong), to the best of our knowledge. Exogenous ketone supplementations are currently studied for potential therapeutic applications, including weight loss, enhanced exercise performance, and the management of neurodegenerative, cardiovascular, and inflammatory conditions [3-5]. We hypothesized that CKM would accurately track BHB and evaluated its performance under sustained intermittent supraphysiological ketosis.

Methods

Study Design

This work is part of a larger study on exogenous ketosis and erythropoiesis (Thomsen et al, unpublished). CKM became available midway through the study and was therefore applied sequentially in the final 7 of the 16 healthy volunteers. Participants were randomized to receive either a ketone ester (KE) drink (500 mg/kg/d) or a placebo (PBO), matched for volume, taste, and viscosity. Over two weeks, drinks were consumed two to three times daily, with half the dose before sleep. Participants were blinded to CKM readings, while investigators were not blinded. We tested the effects of time, treatment, and their interaction on log-transformed BHB area under the curve (AUC) using a linear mixed-effects model and applied polynomial contrasts to assess linear trends.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki II, approved by the regional ethics committee (#1-10-72-221-22), and registered with ClinicalTrials (NCT06053138). Oral and written informed consent was obtained from all participating patients. Participant data were pseudonymized to ensure confidentiality. Participants received financial compensation for their time and participation.

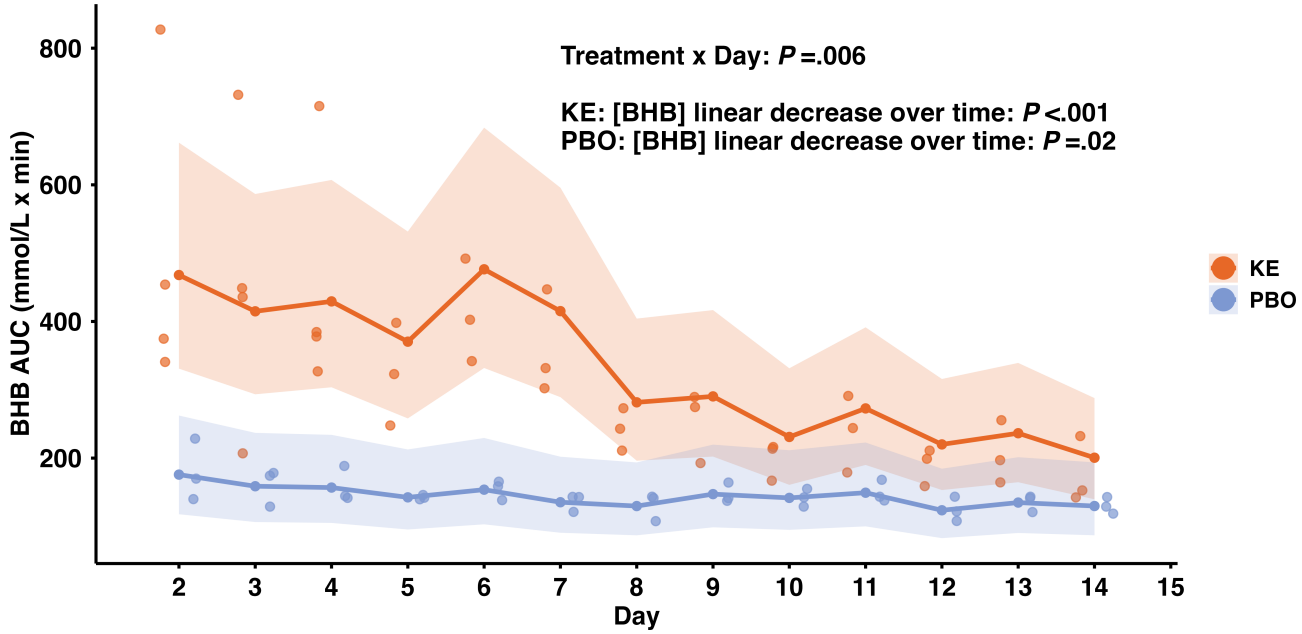
Results

A total of 7 participants wore CKM devices: 4 in the KE group (3 female, 1 male) and 3 in the PBO group (2 female, 1 male). Median age was 41 years (IQR 28–55). One KE participant's sensor detached on day 4 and was not replaced, but CKM

XSL•FO
RenderX

readings until detachment were included in the analyses. BHB AUCs were significantly influenced by both day and treatment, with an interaction effect (*P*=.006). In the KE group, BHB showed a significant linear decrease over 14 days (*P*<.001), and a smaller but significant decline was also observed in the PBO group (*P*=.02). Consequently, group differences diminished, with KE and PBO becoming indistinguishable by the final day (Figure 1).

Figure 1. Day-by-day changes in total BHB area under the curve (AUC) for both the Ketone Ester (KE) group (n=4, orange) and placebo (PBO) group (n=3, blue). Scatter points represent individual AUC measurements for each participant across the 14 study days. Solid lines depict the back-transformed least-square means of BHB concentrations from a mixed-effects model, estimated separately for each day and treatment group, and the shaded regions represent the confidence intervals. BHB: beta-hydroxybutyrate.



Discussion

This study evaluated the performance of a commercially available CKM device during 14 days with intermittent exogenous ketone supplementation. Our findings demonstrate that the CKM detected increases in interstitial BHB concentrations following KE ingestion but revealed a progressive decline in BHB concentrations over the 14-day study period in the KE group, indistinguishable from the PBO group on the last study day. This contrasts with two prior studies in which participants received KE for 14 days before ingesting 25 g KE in a laboratory setting on day 15 [6,7]. In those studies, peaks reached ~2.3 mM at 1 hour and declined to ~0.5 mM at 4 hours, with no evidence of a declining peak BHB concentration following a comparable period of intermittent exogenous ketosis. Importantly, we observed a temporal decline in BHB concentrations also in the placebo group, highly suggesting a ketone-independent physiological or measurement-related drift. Therefore, this raises the possibility of sensor-related limitations. Potential explanations include sensor enzyme degradation, biofouling, temperature effects, compression, or interstitial variability [8]. The underlying sensor principle is not fully disclosed but thought to use a modified electrochemical method reacting selectively with BHB in interstitial fluid. In comparison, an in-development multianalyte sensor using a three-electrode system with NAD^+-dependent β-hydroxybutyrate dehydrogenase and osmium-based redox chemistry has shown stable 14-day performance in 12 healthy, low-carbohydrate-consuming participants [9,10]. A future study is anticipated with interest since it will assess the accuracy of the same device used in our study, SiBio KS1, in subjects following a 14-day ketogenic diet (NCT06420518). Limitations for our study include not comparing the CKM-derived ketone levels with gold standard blood BHB measurements (eg, finger-prick tests), making it difficult to definitively decide if our observations are due to sensor-specific limitations or not. Additionally, the small sample size and statistical power may impact the generalizability of our findings, and it is important to note that the study was not originally designed to evaluate CKM performance.

In conclusion, CKM captured BHB dynamics during exogenous ketosis but revealed a gradual decline in day-to-day BHB AUC over 14 days in both KE and PBO groups, likely reflecting sensor drift rather than physiological adaptation. Larger controlled studies with direct comparison of CKM and blood BHB measurements are needed to confirm accuracy and clinical utility, and must include more than a single batch of CKM devices.

Acknowledgments

The authors express their gratitude to laboratory technologist Trine Larsen at the Department of Clinical Biochemistry in Viborg Regional Hospital, Denmark, for her invaluable technical support.

XSL•FO
RenderX

Funding

The study received financial support from undisclosed research resources within the Medical Diagnostics Center at Regional Hospital Central Jutland, Viborg. Additionally, it was funded by research grants from the Research Funds Pool at Regional Hospital Central Jutland and the Clinical Academic Group for Multimorbidity at the same hospital.

Data Availability

The datasets generated and analyzed in this study are not publicly available because of participant confidentiality and institutional policy restrictions. However, access to the data may be granted upon reasonable request from the corresponding author, subject to the necessary approvals and agreements to ensure data security and adherence to ethical guidelines.

Conflicts of Interest

None declared.

References

1. Moonla C, Reynoso M, Casanova A, et al. Continuous ketone monitoring via wearable microneedle patch platform. ACS Sens 2024 Feb 23;9(2):1004-1013. [doi: 10.1021/acssensors.3c02677] [Medline: 38300831]
2. Kong YW, Morrison D, Lu JC, Lee MH, Jenkins AJ, O'Neal DN. Continuous ketone monitoring: exciting implications for clinical practice. Diabetes Obes Metab 2024 Dec;26 Suppl 7(Suppl 7):47-58. [doi: 10.1111/dom.15921] [Medline: 39314201]
3. Møller N. Ketone body, 3-hydroxybutyrate: minor metabolite - major medical manifestations. J Clin Endocrinol Metab 2020 Sep 1;105(9):dgaa370. [doi: 10.1210/clinem/dgaa370] [Medline: 32525972]
4. Nelson AB, Queathem ED, Puchalska P, Crawford PA. Metabolic messengers: ketone bodies. Nat Metab 2023 Dec;5(12):2062-2074. [doi: 10.1038/s42255-023-00935-3] [Medline: 38092961]
5. Robberechts R, Poffé C. Defining ketone supplementation: the evolving evidence for postexercise ketone supplementation to improve recovery and adaptation to exercise. Am J Physiol Cell Physiol 2024 Jan 1;326(1):C143-C160. [doi: 10.1152/ajpcell.00485.2023] [Medline: 37982172]
6. Berg-Hansen K, Gopalasingam N, Christensen KH, et al. Cardiovascular effects of oral ketone ester treatment in patients with heart failure with reduced ejection fraction: a randomized, controlled, double-blind trial. Circulation 2024 May 7;149(19):1474-1489. [doi: 10.1161/CIRCULATIONAHA.123.067971] [Medline: 38533643]
7. Gopalasingam N, Berg-Hansen K, Christensen KH, et al. Randomized crossover trial of 2-week ketone ester treatment in patients with type 2 diabetes and heart failure with preserved ejection fraction. Circulation 2024 Nov 12;150(20):1570-1583. [doi: 10.1161/CIRCULATIONAHA.124.069732] [Medline: 39162035]
8. Teymourian H, Barfidokht A, Wang J. Electrochemical glucose sensors in diabetes management: an updated review (2010-2020). Chem Soc Rev 2020 Nov 7;49(21):7671-7709. [doi: 10.1039/d0cs00304b] [Medline: 33020790]
9. Alva S, Castorino K, Cho H, Ou J. Feasibility of continuous ketone monitoring in subcutaneous tissue using a ketone sensor. J Diabetes Sci Technol 2021 Jul;15(4):768-774. [doi: 10.1177/19322968211008185] [Medline: 33832353]
10. Abbott's biowearable: one sensor for glucose, ketones. Abbott.com. 2022. URL: https://www.abbott.com/corpnewsroom/strategy-and-strength/abbotts-biowearable-one-sensor-for-glucose-ketones.html [accessed 2025-02-03]

Abbreviations

AUC: area under the curve
KE: ketone ester
BHB: β-hydroxybutyrate
CKM: Continuous ketone monitoring
PBO: placebo
POCT: point-of-care testing

Edited by S Li; submitted 09.Oct.2025; peer-reviewed by D O'Neal, Y Kong; revised version received 22.Dec.2025; accepted 24.Dec.2025; published 13.Jan.2026.

Please cite as:
Kjær SK, Ridder LOR, Svart M, Rittig N, Bjerg LN, Sandfeld-Paulsen B, Thomsen HH
Continuous Ketone Monitoring: Data From a Randomized Controlled Trial
JMIR Diabetes 2026;11:e85548
URL: https://diabetes.jmir.org/2026/1/e85548
doi:10.2196/85548

XSL•FO
RenderX

© Simon K Kjær, Lukas Ochsner Reynaud Ridder, Mads Svart, Nikolaj Rittig, Lise Nørkjær Bjerg, Birgitte Sandfeld-Paulsen, Henrik Holm Thomsen. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 13.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

App-Based Digital Therapeutics Integrating Continuous Glucose Monitoring for Glycemic Control in Type 2 Diabetes: Prospective Observational Cohort Study

Banshi Saboo[1], MD, PhD; Mudit Sabharwal[2], MBBS, FRCP, DFM, PG Dip Diab; Mihir Gharia[3], MBBS; Devina Aswal[3], MBBS; Parth Modi[3], MBBS, MD; Twinkle Maheshwari[3], PharmD; Jaymin Parikh[3], PhD; Simpel Saanyal[3], BDS; Jaikishan Agrawal[1], PhD

[1]Diabetes Care and Hormone Clinic, Dia Care Research, Ahmedabad, Gujarat, India

[2]Department of Diabetes and Endocrinology, Dharma Diabetes and Metabolic Clinics, Delhi, India

[3]Department of Medical Affairs, Tatvacare, Purple H 202, Maple County 2, Sindhu Bhavan Road Thaltej, Ahmedabad, Gujarat, India

Corresponding Author:
Mihir Gharia, MBBS
Department of Medical Affairs, Tatvacare, Purple H 202, Maple County 2, Sindhu Bhavan Road Thaltej, Ahmedabad, Gujarat, India

Abstract

Background: Digital therapeutics integrating continuous glucose monitoring (CGM) with personalized lifestyle coaching can enhance glycemic control in individuals with type 2 diabetes mellitus (T2DM). However, real-world evidence evaluating such multicomponent interventions—combining CGM with nutrition coaching, physiotherapy, and cognitive behavioral therapy within a unified digital platform—remains limited in South Asian populations.

Objective: This study aimed to evaluate the effectiveness of an app-based lifestyle-integrated intervention program (Glycemic Lifestyle Intervention in Diabetes Empowerment–CGM; GLIDE-C) on glycemic, anthropometric, lipid profile, quality of life (QoL), and behavioral outcomes, as well as outcomes related to HEOR (health economics and outcomes research), in adults with mildly uncontrolled T2DM.

Methods: A prospective observational cohort study was conducted among adults aged ≥18 years with confirmed T2DM and baseline hemoglobin A_{1c} (HbA_{1c}) between 7.5% and 9.5%. The intervention was delivered via the Goodflip mobile app, including a 15-day CGM period and a structured analysis of glucose variations, among other clinical assessments, with follow-up on days 15, 30, and 60. The digital intervention combined CGM-guided personalized diet, exercise, and cognitive behavioral therapy plans delivered through the mobile app, supported by multidisciplinary coaching. Primary outcomes were changes in HbA_{1c}, fasting plasma glucose, postprandial glucose, and CGM metrics; secondary outcomes included lipid profile, anthropometric indices, QoL, and behavioral determinants of glycemic improvement.

Results: Eighteen participants (mean age 49.45, SD 10.91 years; n=11, 61.1% male) completed the program. Mean HbA_{1c} decreased by 0.59% (P=.05), with postprandial glucose decreasing by 25.42 mg/dL (P=.06) and time in range increasing by 9.98%. Modest reductions, consistent with the short intervention duration, occurred in low-density lipoprotein cholesterol (−8.38 mg/dL), body fat (−1.13 kg), and waist circumference (−1.56 cm). QoL improved significantly for sleep quality (P=.04). Improvements were also observed for motivation (P=.06), stress management (P=.08), and mood (P=.12), although these changes were not statistically significant. Behavioral analysis demonstrated strong associations with glycemic outcomes; all participants who reduced their consumption of energy-dense, nutrient-poor foods experienced improvements in HbA_{1c}, whereas those who did not reduce energy-dense, nutrient-poor intake showed less favorable outcomes.

Conclusions: The novel GLIDE-C program demonstrated clinically meaningful improvements in glycemic control, behavioral outcomes, and QoL in adults with T2DM, without pharmacological modification. These findings strengthen the role of integrated, CGM-guided digital therapeutic platforms as an effective adjunct to the conventional standard of care.

Trial Registration: Clinical Trials Registry-India CTRI/2024/08/071850; https://ctri.nic.in/Clinicaltrials/pmaindet2.php?EncHid=MTEyMzA2&Enc=&userName=

(JMIR Diabetes 2026;11:e86651) doi:10.2196/86651

KEYWORDS

type 2 diabetes mellitus; continuous glucose monitoring; digital therapeutics; lifestyle intervention; cognitive behavioral therapy

XSL•FO
RenderX

Introduction

Type 2 diabetes mellitus (T2DM) constitutes one of the most consequential noncommunicable disease epidemics of the 21st century. The International Diabetes Federation Diabetes Atlas (11th edition) estimates that 589 million adults worldwide were living with diabetes in 2024, a figure projected to reach 853 million by 2050, with the largest relative increases expected in low- and middle-income countries [1]. The landmark national Indian Council of Medical Research-India Diabetes cross-sectional study reported a diabetes prevalence of 11.4% among Indian adults, with an estimated 101 million people affected and a further 136 million living with prediabetes across the country [2]. South Asian individuals are phenotypically predisposed to insulin resistance at lower BMIs, develop T2DM at a younger age, and carry a higher burden of complications relative to their Western counterparts [3]. Effective glycemic management is therefore of paramount clinical importance. Glycemic monitoring relies on hemoglobin A_{1c} (HbA_{1c}) and self-monitoring of blood glucose (SMBG). While HbA_{1c} remains the principal marker of long-term glycemic control, it is insensitive to intraday glucose variability, postprandial excursions, and nocturnal hypoglycemia. Similarly, SMBG provides only episodic data and is highly dependent on patient adherence and technique [4]. Continuous glucose monitoring (CGM), a more advanced glucose monitoring technology, enables real-time assessment of interstitial glucose levels and provides a more comprehensive view of glycemic patterns and variability [4]. However, translating these insights to a sustained behavioral change requires a structured clinical, nutritional, and psychological support framework.

A 2024 systematic review and meta-analysis demonstrated that real-time CGM and intermittently scanned CGM both produce clinically meaningful HbA_{1c} reductions compared with SMBG across insulin and noninsulin treatment regimens [5]. Similar results have been demonstrated in larger cohorts (n=74,679), with reductions in all-cause hospitalizations and acute diabetes-related emergencies [6]. However, certain limitations persist, such as the results possibly being attenuated without behavioral and lifestyle components [7]. The concurrent evidence base for digital therapeutics and app-based interventions in T2DM is characterized by heterogeneity in both design and outcomes. Most interventions focus on behavioral therapy, nutrition guidance, physical activity, or various combinations thereof and are typically not comprehensive [7-9]. In the Indian context, real-world digital therapeutics programs have demonstrated improvements in glycemic and weight-related outcomes among people with T2DM [10,11]. A diabetes CGM digital therapeutics program, closely analogous to prior real-world evaluation, demonstrated a reduction in all glycemic parameters but lacked a structured incorporation of physiotherapy, cognitive behavioral therapy (CBT), or multidomain quality of life (QoL) assessment [10]. Collectively, these limitations, due to single-component or partially integrated interventions, can achieve improvements in isolation; however, behavior change sustainability and patient-centered outcomes remain inconsistently addressed [12,13].

Existing literature reveals critical gaps. There is a lack of integrated multidisciplinary digital therapeutics combining CGM-guided glycemia, personalized nutrition, exercise prescription, and CBT. Real-world evidence remains limited beyond controlled trials. South Asian or Indian populations are underrepresented despite their high T2DM burden [3,11]. No studies evaluate CGM metrics, HbA_{1c}, anthropometrics, behavior, QoL, and health economics in a single program.

To address these gaps, this study introduces a novel, fully integrated digital therapeutics model, GLIDE-C, which uniquely combines CGM-guided glycemic monitoring with personalized nutrition, structured physiotherapy, and CBT within a single digital ecosystem. Unlike prior interventions that evaluate isolated components or limited outcomes, this study simultaneously assesses multidomain end points, including glycemic metrics such as HbA_{1c} and CGM parameters, anthropometrics, behavioral changes, QoL, and exploratory health economic outcomes in a real-world Indian population. The aim of this study was to evaluate the real-world effectiveness of the GLIDE-C program over 90 days among Indian adults with T2DM.

Methods

Study Design and Setting

This prospective, observational cohort study was conducted at DiaCare Research, Ahmedabad, between June 2024 and December 2024 using deidentified program data from adults enrolled in the GLIDE-C program, a structured 90-day digital therapeutics intervention delivered via the Goodflip mobile app. This observational study was supported by a multidisciplinary health coaching team.

Participants

Eligible participants were adults aged ≥18 years with a confirmed diagnosis of T2DM and a baseline HbA_{1c} between 7.5% and 9.5%. All participants were receiving ongoing standard-of-care antidiabetic therapy as prescribed by their treating physician. Additional eligibility criteria included ownership of a smartphone with internet access and sufficient English literacy to engage with the program content and complete study questionnaires. Participants were excluded if they were currently or had recently been involved in other structured lifestyle intervention programs, were taking weight-loss supplements or pharmacotherapy, had no recent intensification or initiation of new antidiabetic medications in the month preceding enrollment, were pregnant or breastfeeding within the past 3 months, had undergone recent surgery, or had advanced chronic kidney disease (stage 4 or 5) or liver cirrhosis.

GLIDE-C Program

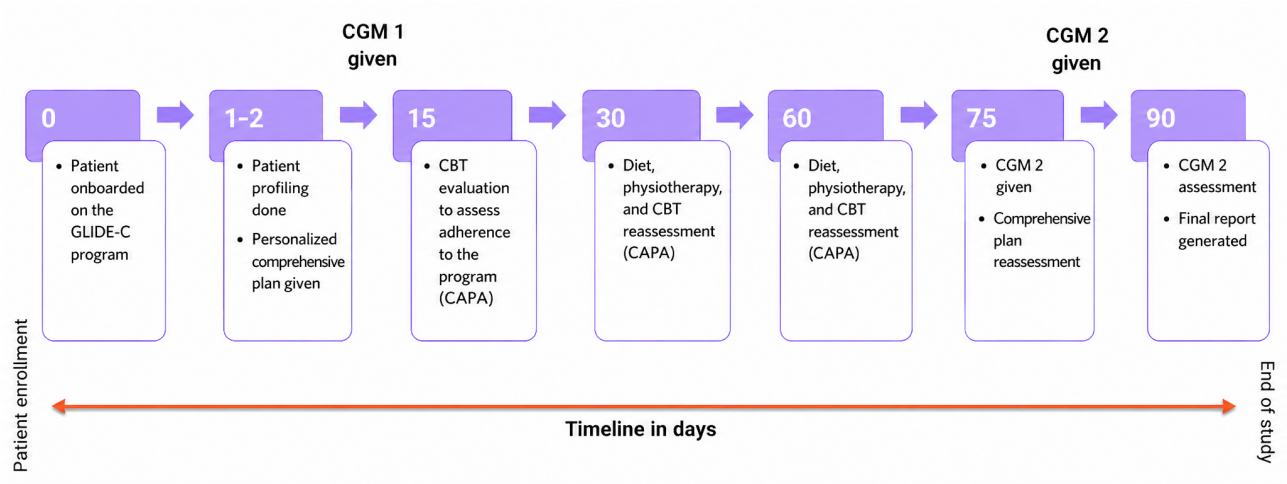
GLIDE-C is a multicomponent integrated digital therapeutics program designed to support glycemic control through personalized lifestyle modification. The intervention was delivered through the Goodflip mobile app (version 5.2.23; iOS and Android), which integrates CGM data, Internet of Things (IoT) devices, educational content, behavioral therapy modules, and 2-way communication with health coaches. Upon

XSL•FO
RenderX

enrollment, participants completed a baseline consultation where demographic characteristics, dietary habits, physical activity levels, and QoL measures were captured. QoL domains were assessed using custom, program-developed scales and not externally validated instruments. Following this assessment, each participant was assigned a dedicated multidisciplinary team comprising a dietitian, physiotherapist, and CBT.

The app visualized key CGM metrics, time in range (TIR), time above range (TAR), and time below range (TBR) and generated personalized glucose alerts adjustable to clinical thresholds. At the end of the initial CGM period, a structured root cause analysis (for diet, behavior, and exercise) was conducted, followed by modifications to the intervention plan as needed. The nutrition and physiotherapy coaches guided the patients with personalized plans, while the CBT interventions focused on addressing motivational barriers, stress management, and emotional eating using mindfulness training, relaxation techniques, and cognitive restructuring. Participants' journey through the program is depicted in Figure 1.

Figure 1. Participants' journey through the 90-day GLIDE-C program. CAPA: corrective and preventive actions; CBT: cognitive behavioral therapy; CGM: continuous glucose monitoring.



Outcomes

The primary outcomes were changes from baseline to day 90 in HbA_{1c}, fasting plasma glucose, postprandial glucose, CGM-derived measures (TIR, TAR, and TBR), anthropometric parameters (including weight, BMI, waist circumference, and body fat percentage), and lipid profile measures. Secondary outcomes included changes in QoL parameters and exploratory health economics and outcomes research (HEOR) measured through reduction in insulin dose and extension in follow-up visits calibrated at the end of the study.

Data Analysis

The deidentified data for all the participants were taken from the Goodflip mobile app. Continuous variables were summarized as mean (SD). Normality of distribution was assessed using the Shapiro-Wilk test. For paired baseline and postintervention comparisons, the paired *t* test (2-tailed) was applied to normally distributed variables, while the Wilcoxon signed-rank test was used for nonnormally distributed variables. Percentage changes from baseline were calculated for each outcome. Associations between behavioral changes (nutrient-rich food intake; energy-dense, nutrient-poor [EDNP] food intake; and weekly exercise) and HbA_{1c} improvement were explored descriptively. All analyses were performed using SPSS (version 27.0; IBM Corp), with 2-tailed P values <.05 considered statistically significant.

Ethical Considerations

The study protocol was reviewed and approved by the institutional ethics committee of Aatman Hospital, Ahmedabad, India (EC/NEW/INST/2021/11802). The study was prospectively registered with the Clinical Trials Registry-India (CTRI/2024/08/071850). All study procedures adhered to the principles outlined in the Declaration of Helsinki and its subsequent amendments. Electronic informed consent was obtained from each participant before enrollment; participants were informed of the study purpose, the voluntary nature of participation, and their right to withdraw at any time without any change to their ongoing standard of care. The data collected were deidentified prior to analysis, with restricted access to authorized personnel only.

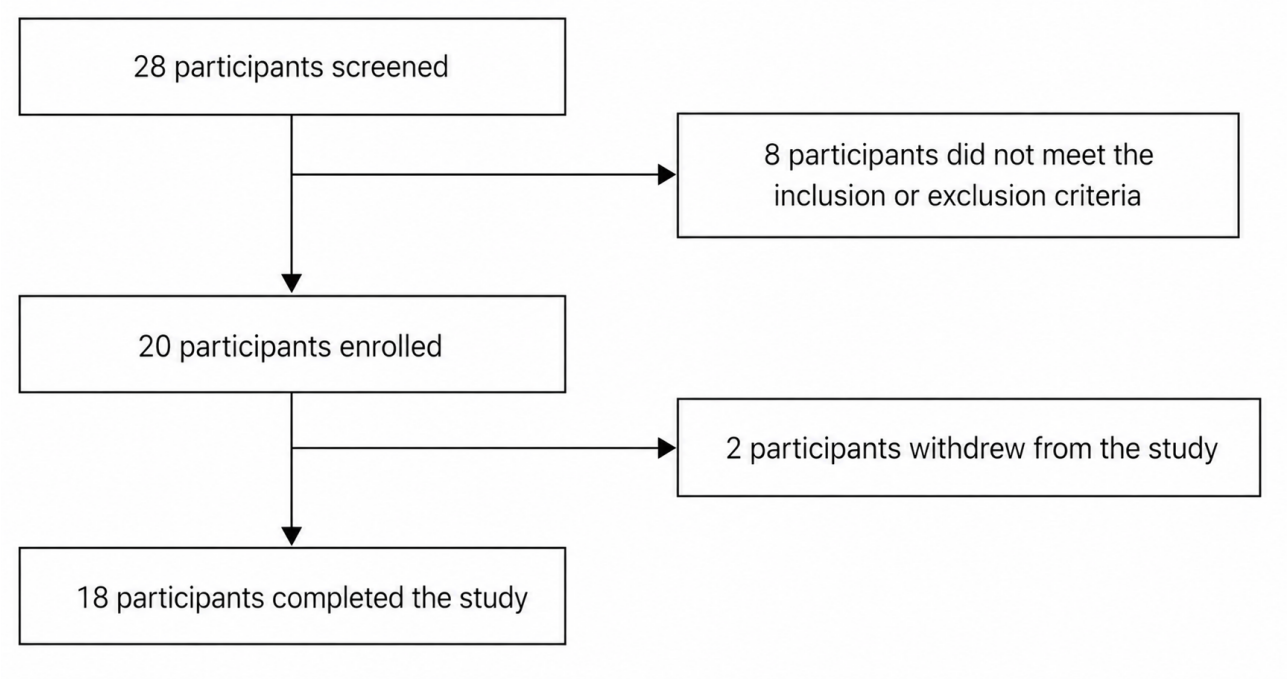
Results

Participant Characteristics

A total of 20 participants were enrolled after screening 28 participants; 18 participants completed the 90-day program and were included in the final analysis (90% completion rate). The 2 noncompleters withdrew voluntarily for reasons unrelated to the intervention or adverse events; no protocol-mandated withdrawals occurred (Figure 2). Baseline anthropometric, glycemic, lipid, and CGM parameters are summarized in Table 1.

XSL•FO
RenderX

Figure 2. Participant flow.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of participants (N=18).

Variable	Values
Age (years), mean (SD)	49.45 (10.91)
Sex, n (%)	
Female	7 (38.9)
Male	11 (61.1)
Employment status, n (%)	
Student	0 (0)
Housewife	3 (16.7)
Working	13 (72.2)
Retired	2 (11.1)
Diet type, n (%)	
Vegetarian	10 (55.6)
Nonvegetarian	8 (44.4)
Body fat mass (kg), mean (SD)	35.39 (6.38)
Weight (kg), mean (SD)	82.06 (9.66)
Skeletal muscle mass (kg), mean (SD)	25.45 (4.16)
Waist circumference (cm), mean (SD)	105.84 (7.64)
HbA_{1c}[a] (%), mean (SD)	8.10 (0.41)
Fasting plasma glucose (mg/dL), mean (SD)	143.68 (28.54)
Postprandial glucose (mg/dL), mean (SD)	194.16 (28.90)
Total cholesterol (mg/dL), mean (SD)	138.61 (30.43)
Triglycerides (mg/dL), mean (SD)	174.95 (72.14)
HDL[b] cholesterol (mg/dL), mean (SD)	41.16 (11.11)
LDL[c] cholesterol (mg/dL), mean (SD)	74.47 (25.07)
VLDL[d] cholesterol (mg/dL), mean (SD)	45.80 (28.36)
TIR[e] (%), mean (SD)	74.37 (12.14)
TBR[f] (%), mean (SD)	3.47 (3.29)
TAR[g] (%), mean (SD)	22.16 (13.43)

[a]HbA_{1c}: hemoglobin A_{1c}.
[b]HDL: high-density lipoprotein.
[c]LDL: low-density lipoprotein.
[d]VLDL: very low-density lipoprotein.
[e]TIR: time in range.
[f]TBR: time below range.
[g]TAR: time above range.

Primary Outcomes

Glycemic Control

At 90 days, mean HbA_{1c} decreased by 0.59 percentage points (95% CI 0.00-1.18; *P*=.05), representing a relative reduction of 7.3%. Fasting plasma glucose declined by 10.94 mg/dL (95% CI −11.0 to 32.9; *P*=.32; the relative percentage change from baseline to 90 days was −7.6%) and postprandial glucose by 25.42 mg/dL (95% CI −1.5 to 52.3), with a trend toward statistical significance (*P*=.06; −13.1%). CGM metrics showed a 9.98% relative increase in TIR from 74.37% to 81.79% (*P*=.16), a 32.3% reduction in TAR from 22.16% to 15.00% (*P*=.20), and no significant change in TBR (*P*=.84; Table 2).

XSL•FO
RenderX

Table . Changes in glycemic and continuous glucose monitoring parameters from baseline to 90 days (N=18).

Parameter	Baseline, mean (SD)	90 days, mean (SD)	Absolute change (95% CI)	Percentage change	*P* value
HbA_{1c}[a] (%)	8.10 (0.41)	7.51 (1.17)	−0.59 (0.00 to −1.18)	−7.28	.05[b]
FPG[c] (mg/dL)	143.68 (28.54)	132.74 (31.98)	−10.94 (−11.0 to 32.9)	−7.61	.32
PPG[d] (mg/dL)	194.16 (28.90)	168.74 (42.34)	−25.42 (−1.5 to 52.3)	−13.09	.06
TIR[e] (%)	74.37 (12.14)	81.79 (16.89)	+7.42 (−3.4 to 18.2)	+9.98	.16
TBR[f] (%)	3.47 (3.29)	3.74 (3.38)	+0.27 (−2.6 to 3.1)	+7.78	.84
TAR[g] (%)	22.16 (13.43)	15.00 (17.72)	−7.16 (−4.0 to 18.3)	−32.31	.20

[a]HbA_{1C}: hemoglobin A_{1c}.
[b]Statistically significant.
[c]FPG: fasting plasma glucose.
[d]PPG: postprandial glucose.
[e]TIR: time in range.
[f]TBR: time below range.
[g]TAR: time above range.

Anthropometric Measures

Anthropometric outcomes indicated small reductions in adiposity while maintaining lean muscle mass. Body fat mass fell by 1.13 kg (−3.19%), waist circumference by 1.56 cm (−1.47%), and weight by 0.67 kg (−0.82%). Skeletal muscle mass increased marginally (+0.33 kg; +1.3%). None of these changes were statistically significant (Table 3).

Table . Changes in anthropometric measures from baseline to 90 days (N=18).

Parameter	Baseline, mean (SD)	90 days, mean (SD)	Absolute change (95% CI)	Percentage change	*P* value
Body fat mass (kg)	35.39 (6.38)	34.26 (7.82)	−1.13 (−4.1 to 6.3)	−3.19	.66
Skeletal muscle mass (kg)	25.45 (4.16)	25.78 (5.06)	+0.33 (−3.2 to 2.5)	+1.30	.84
Weight (kg)	82.06 (9.66)	81.39 (13.21)	−0.67 (−7.5 to 6.2)	−0.82	.88
Waist circumference (cm)	105.84 (7.64)	104.28 (9.95)	−1.56 (−4.6 to 7.7)	−1.47	.62

Lipid Profile

Lipid parameters showed modest but statistically substantial changes over 90 days. Low-density lipoprotein cholesterol decreased by 8.38 mg/dL (−11.25%), total cholesterol by 10.14 mg/dL (−7.32%), and very low-density lipoprotein by 9.06 mg/dL (−19.78%). High-density lipoprotein cholesterol remained stable, and triglycerides increased slightly by 8.73 mg/dL (+4.99%; Table 4).

XSL•FO
RenderX

Table . Changes in lipid profile from baseline to 90 days (N=18).

Parameter	Baseline, mean (SD)	90 days, mean (SD)	Absolute change (95% CI)	Percentage change	*P* value
Total cholesterol (mg/dL)	138.61 (30.43)	128.47 (25.91)	−10.14 (−12.0 to 32.3)	−7.32	.36
Triglycerides (mg/dL)	174.95 (72.14)	183.68 (59.27)	+8.73 (−41.2 to 23.7)	+4.99	.72
HDL[a] cholesterol (mg/dL)	41.16 (11.11)	41.21 (9.26)	+0.05 (−4.9 to 4.8)	+0.12	.99
LDL[b] cholesterol (mg/dL)	74.47 (25.07)	66.09 (23.44)	−8.38 (−10.7 to 27.5)	−11.25	.38
VLDL[c] cholesterol (mg/dL)	45.80 (28.36)	36.74 (20.88)	−9.06 (−12.0 to 30.1)	−19.78	.48

[a]HDL: high-density lipoprotein.

[b]LDL: low-density lipoprotein.

[c]VLDL: very low-density lipoprotein.

Secondary Outcomes

Over the 90-day program, participants reported improvements across multiple QoL domains. Sleep quality increased by 18.9 points (95% CI 61.11-80.00; *P*=.04), motivation by 13.06 points (95% CI 59.44-72.50; *P*=.06), stress management by 12.78 points (95% CI 65.55-78.33; *P*=.08), and mood by 10.55 points (95% CI 66.11-76.66; *P*=.12). A composite CBT score, encompassing 6 domains (mood, motivation, time management, stress management, sleep, and awareness), improved in all participants, with improvements ranging from 4.3% to 80%. Half of the cohort achieved a ≥20% increase, indicating a broad enhancement in psychosocial well-being linked to personalized lifestyle coaching and CGM-driven feedback (Table 5).

Improvement in HbA_{1c} was observed in 15 (83.3%) of the 18 participants. Among those who increased their intake of nutrient-rich foods, 92.9% (n=17) achieved an improvement in HbA_{1c} compared with 50% (n=9) of participants who did not make such dietary changes. All participants who reduced their consumption of EDNP foods demonstrated improvement in HbA_{1c}, whereas none of the participants who maintained or increased EDNP intake showed any improvement. Similarly, 90% (n=16) of participants engaging in regular weekly exercise achieved HbA_{1c} improvement compared with 75% (n=13) of those who were not exercising regularly (Table 6).

At 3 months, HbA_{1c} reduction was the most frequently observed HEOR outcome, occurring in 75% (13/18) of participants, including 2 participants who were receiving insulin therapy at baseline. However, nearly 25% (4/18) of participants were nonresponders and did not show improvement in glycemic parameters. In comparison, follow-up interval extension and weight reduction were observed in 25% (4/18) and 61% (11/18) of participants, respectively. Additionally, both participants who were receiving insulin therapy at baseline showed a reduction in insulin doses at the end of the study.

Table . Changes in quality of life scores from baseline to 90 days.

Domain	Baseline, mean (SD)	90 days, mean (SD)	Absolute change	Percentage change	*P* value
Sleep quality	61.11 (26.1)	78.46 (10.7)	+18.89	+30.9	.04
Motivation	59.44 (21.0)	72.86 (10.7)	+13.06	+22.0	.06
Stress management	65.55 (24.8)	79.23 (10.4)	+12.78	+19.5	.08
Mood	66.11 (22.7)	76.92 (12.5)	+10.55	+15.9	.12

XSL•FO
RenderX

Table . Relationship between behavioral factors and hemoglobin A_{1c} (HbA_{1c}) improvement (N=18).

Behavioral factor	HbA_{1c} improved, n/N (%)	HbA_{1c} not improved, n/N (%)	Total, n/N (%)
Increased NRF[a] intake	13/14 (92.9)	1/14 (7.1)	14/18 (77.8)
No increase in NRF intake	2/4 (50)	2/4 (50)	4/18 (22.2)
Reduced EDNP[b] intake	15/15 (100)	0/15 (0)	15/18 (83.3)
No reduction in EDNP intake	0/3 (0)	3/3 (100)	3/18 (16.7)
Regular weekly exercise	9/10 (90)	1/10 (10)	10/18 (55.6)
Not regular weekly exercise	6/8 (75)	2/8 (25)	8/18 (44.4)

[a]NRF: nutrient-rich foods.

[b]EDNP: energy-dense, nutrient-poor.

Discussion

Principal Findings

The GLIDE-C program is the first to evaluate a CGM-enabled, multidisciplinary digital therapeutic integrating nutrition, physiotherapy, and CBT with simultaneous assessment of clinical, behavioral, and QoL outcomes in a real-world Indian T2DM cohort, a population that carries a disproportionate share of global diabetes burden that remains underrepresented in digital therapeutics research.

We observed a 0.59-point reduction in HbA_{1c}, an almost 10% rise in TIR, and a 32% fall in TAR, with 83% (15/18) of participants improving their glycemic control. These metabolic gains were accompanied by improvements in mood, motivation, sleep, stress, and anthropometric parameters, without loss of muscle mass.

CGM transforms glucose data from episodic snapshots into dynamic profiles, allowing users to see the glycemic impact of specific foods, physical activities, and stressors. This immediate feedback supports “closed-loop” learning, reinforcing beneficial behaviors and discouraging detrimental ones [12]. In contrast, traditional fingerstick self-monitoring provides limited context and often fails to motivate sustained change. CGM provides visual feedback that reduces the cognitive burden associated with interpreting the effects of dietary choices and physical activity, thereby lowering behavioral barriers to effective self-management [4,14]. Positive trends in the app serve as positive reinforcements and create an aspiration for self-comparison [7,12].

Reviews note that the high upfront cost of CGM is often offset by fewer complications and hospitalizations, as it reduces both hyperglycemia and hypoglycemia [14]. Randomized and real-world studies show that CGM reduces HbA_{1c} by 0.25% to 3% and increases TIR by 15% to 34%, while also reducing hypoglycemia and improving treatment satisfaction [14]. In the BlueStar real-time CGM program, adults with poor baseline control increased TIR by 15 percentage points and reduced hyperglycemia by 14 percentage points over 3 months, despite limited weekly engagement; participants attributed their success to “teachable moments” from real-time glucose feedback and artificial intelligence–driven coaching [12]. A 7.42-point rise in TIR in this study is comparable, given the shorter duration and relatively good baseline control. Studies combining CGM with digital coaching report similar or greater effects: January AI and January V2 (artificial intelligence–supported CGM mobile apps) increased TIR by 10% and achieved modest weight loss in cohorts tracked for 14 to 28 days [15,16]. The digital twin program, which leverages machine learning and IoT sensors to deliver year-long personalized nutrition and activity plans, achieved a 1.8% HbA_{1c} reduction, weight loss of 4.8 kg, and a reduction in antidiabetic medication use from 1.9 to 0.5 drugs per person; nearly 89% of participants attained HbA_{1c} <7% [17]. Digital applications that focus on a single behavioral domain yield a more limited glycemic effect. In a 12-week randomized trial of a smartphone personal health record app with motivational SMS text messages, HbA_{1c} reduced from 6.9% to 6.7% in the intervention arm but remained similar in the control arm, and the primary outcome, step count, did not differ [18]. A 6-month trial of the Emperra GmbH E-Health Technologies digital application (largely for insulin-treated patients) reported a 0.48% reduction in HbA_{1c} vs a 0.28% difference relative to standard care [19]. More recently, the Vitadio mobile health app, which combines educational content, self-monitoring, and virtual coaching, was associated with a greater reduction in HbA1c than standard care (−0.8% vs −0.3%), comparable to the findings of this study, as well as significant reductions in weight and BMI and improvements in self-care practices [20].

Individuals who increase consumption of nutrient-rich foods typically achieve greater HbA_{1c} improvement, with average reductions of about 0.3% to 0.4%, compared with those who do not, while those who reduce EDNP foods consistently show improvements of around 0.5% [21,22]. These findings align with digital nutrition interventions such as Vitadio and in our GLIDE-C study, where significant improvements in diet quality drove HbA_{1c} reduction and weight loss [20]. Regular exercise is also associated with HbA_{1c} improvement; this mirrors data from physical activity apps, where participants with low baseline activity achieved greater step count and HbA_{1c} improvements [18].

The individual plan, including CBT for all participants, supports and fosters autonomous motivation, which may empower patients in the long term [8]. These findings suggest that sustained engagement with data-driven programs can yield durable metabolic benefits through behavior modification

XSL•FO
RenderX

[16,17], leading to improvements in health-related QoL and, in some cases, medication de-escalation as clinical outcomes improve, as observed in this study [9,13]. Additionally, in this study, reductions in HbA_{1c} and body weight were associated with reduction in insulin doses and longer intervals between follow-up visits, collectively reflecting an overall improvement in HEOR outcomes.

This study implements the synergistic effects of diet quality, exercise, CBT, and IoT devices through a single digital platform unifying all the interventions from available literature. South Asian individuals characterized by insulin resistance at lower BMI thresholds and earlier onset of T2DM may respond differently to lifestyle-based digital interventions compared with Western populations [2,3,11]. The findings of this study therefore carry specific translational relevance for health systems in India and the broader South Asian region, where scalable, low-infrastructure digital solutions are urgently needed [1,10]. This real-world observational cohort study provides preliminary evidence for a CGM-integrated digital therapeutic application. The single-arm design and the current sample size are consistent with published feasibility studies and sufficiently detect clinically meaningful improvements across glycemic and behavioral domains [7,8,10,11,15]. However, the absence of a randomized control group precludes causal inference and may reflect insufficient statistical power for secondary end points, while regression to the mean cannot be excluded. The time frame, though standard in digital therapeutics trials, captured early program effects; longer follow-up in future studies will enable us to characterize the sustainability of outcomes.

Conclusions

The novel GLIDE-C program, integrating CGM with personalized nutrition, physiotherapy, and CBT, achieved meaningful improvements in glycemic control, QoL, and health behaviors in adults with T2DM. Real-time data review and tailored coaching fostered dietary and activity changes that were strongly associated with reductions in HbA_{1c}. These findings highlight the potential of holistic, app-based digital therapeutics to complement conventional diabetes care and support sustainable lifestyle modification in real-world settings.

Acknowledgments

The authors confirm that no generative artificial intelligence tools, including but not limited to ChatGPT, Claude, Gemini, or other similar large language models, were used in the preparation of this manuscript, including writing, editing, data analysis, or figure generation.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Data Availability

This study used deidentified data from a proprietary digital therapeutics platform. Access to the full dataset is restricted due to participant confidentiality agreements and commercial data governance requirements. However, aggregate data supporting the findings of this study are available from the corresponding author upon reasonable request. Researchers wishing to access the data may contact the corresponding author with a brief description of the intended use.

Authors' Contributions

Conceptualization: BS (lead), MS (equal)
Data curation: MG, DA, PM, TM, JP, SS
Formal analysis: PM (supporting)
Investigation: MS, JA
Methodology: MS (lead), MG (supporting)
Project administration: BS (lead), MS (equal), MG (supporting)
Resources: JA
Supervision: BS
Validation: MG (lead), PM (supporting)
Visualization: TM (lead)
Writing – original draft: MG (lead), DA (supporting), TM (supporting)
Writing – review and editing: BS (lead), MS (supporting), TM (supporting), JP (supporting)

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. IDF Diabetes Atlas, 11th edition: International Diabetes Federation; 2025. URL: https://diabetesatlas.org/ [accessed 2026-06-21]
2. Anjana RM, Unnikrishnan R, Deepa M, et al. Metabolic non-communicable disease health report of India: the ICMR-INDIAB national cross-sectional study (ICMR-INDIAB-17). Lancet Diabetes Endocrinol 2023 Jul;11(7):474-489. [doi: 10.1016/S2213-8587(23)00119-5] [Medline: 37301218]
3. Nanditha A, Susairaj P, Satheesh K, Raghavan A, Snehalatha C, Ramachandran A. The rising prevalence of type 2 diabetes among the youth in southern India-an ancillary analysis of the Secular TRends in DiabEtes in India (STRiDE-I) study. J Diabetes 2024 Jul;16(7):e13576. [doi: 10.1111/1753-0407.13576] [Medline: 38923743]
4. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023 Jan;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9] [Medline: 36493795]
5. Jancev M, Vissers T, Visseren FLJ, et al. Continuous glucose monitoring in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetologia 2024 May;67(5):798-810. [doi: 10.1007/s00125-024-06107-6] [Medline: 38363342]
6. Garg SK, Hirsch IB, Repetto E, et al. Impact of continuous glucose monitoring on hospitalizations and glucose control in people with type 2 diabetes: real-world analysis. Diabetes Obes Metab 2024 Nov;26(11):5202-5210. [doi: 10.1111/dom.15866] [Medline: 39263872]
7. Bodner E, Roth L, Wiencke K, Bischoff C, Schwarz PE. Effect of multimodal app-based interventions on glycemic control in patients with type 2 diabetes: systematic review and meta-analysis. J Med Internet Res 2025 Jan 24;27:e54324. [doi: 10.2196/54324] [Medline: 39854703]
8. Hsia J, Guthrie NL, Lupinacci P, et al. Randomized, controlled trial of a digital behavioral therapeutic application to improve glycemic control in adults with type 2 diabetes. Diabetes Care 2022 Dec 1;45(12):2976-2981. [doi: 10.2337/dc22-1099] [Medline: 36181554]
9. Lee YB, Kim G, Jun JE, et al. An integrated digital health care platform for diabetes management with AI-based dietary management: 48-week results from a randomized controlled trial. Diabetes Care 2023 May 1;46(5):959-966. [doi: 10.2337/dc22-1929] [Medline: 36821833]
10. Joshi S, Verma R, Lathia T, et al. Fitterfly Diabetes CGM digital therapeutics program for glycemic control and weight management in people with type 2 diabetes mellitus: real-world effectiveness evaluation. JMIR Diabetes 2023 May 3;8:e43292. [doi: 10.2196/43292] [Medline: 37133922]
11. Chawla R, Jaggi S, Gupta A, Bantwal G, Patil S. Clinical utility of a digital therapeutic intervention in Indian patients with type 2 diabetes mellitus: 12-week prospective single-arm intervention study. JMIR Diabetes 2022 Oct 31;7(4):e41401. [doi: 10.2196/41401] [Medline: 36226840]
12. Kumbara AB, Iyer AK, Green CR, et al. Impact of a combined continuous glucose monitoring-digital health solution on glucose metrics and self-management behavior for adults with type 2 diabetes: real-world, observational study. JMIR Diabetes 2023 Sep 11;8:e47638. [doi: 10.2196/47638] [Medline: 37590491]
13. Oser TK, Litchman ML, Allen NA, et al. Personal continuous glucose monitoring use among adults with type 2 diabetes: clinical efficacy and economic impacts. Curr Diab Rep 2021 Dec 9;21(11):49. [doi: 10.1007/s11892-021-01408-1] [Medline: 34882273]
14. Kwon SY, Moon JS. Advances in continuous glucose monitoring: clinical applications. Endocrinol Metab (Seoul) 2025 Apr;40(2):161-173. [doi: 10.3803/EnM.2025.2370] [Medline: 40195726]
15. Zahedani AD, McLaughlin T, Veluvali A, et al. Digital health application integrating wearable data and behavioral patterns improves metabolic health. NPJ Digit Med 2023 Nov 25;6(1):216. [doi: 10.1038/s41746-023-00956-y] [Medline: 38001287]
16. Veluvali A, Dehghani Zahedani A, Hosseinian A, et al. Impact of digital health interventions on glycemic control and weight management. NPJ Digit Med 2025 Jan 9;8(1):20. [doi: 10.1038/s41746-025-01430-7] [Medline: 39789102]
17. Shamanna P, Erukulapati RS, Shukla A, et al. One-year outcomes of a digital twin intervention for type 2 diabetes: a retrospective real-world study. Sci Rep 2024 Oct 26;14(1):25478. [doi: 10.1038/s41598-024-76584-7] [Medline: 39461977]
18. Kim G, Kim S, Lee YB, Jin SM, Hur KY, Kim JH. A randomized controlled trial of an app-based intervention on physical activity and glycemic control in people with type 2 diabetes. BMC Med 2024 May 1;22(1):185. [doi: 10.1186/s12916-024-03408-w] [Medline: 38693528]
19. Roth L, Steckhan N, Schwarz PE. Impact of a digital application on HbA1c levels in people with diabetes: a randomized controlled trial. Front Digit Health 2025;7:1544668. [doi: 10.3389/fdgth.2025.1544668] [Medline: 40661654]
20. Bretschneider MP, Kolasińska AB, Šomvárska L, Klásek J, Mareš J, Schwarz PE. Evaluation of the impact of mobile health app Vitadio in patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2025 May 9;27:e68648. [doi: 10.2196/68648] [Medline: 40344662]
21. Yen TS, Htet MK, Lukito W, et al. Increased vegetable intake improves glycaemic control in adults with type 2 diabetes mellitus: a clustered randomised clinical trial among Indonesian white-collar workers. J Nutr Sci 2022;11:e49. [doi: 10.1017/jns.2022.41] [Medline: 35836691]
22. Heidari Seyedmahalleh M, Nasli-Esfahani E, Zeinalabedini M, Azadbakht L. Association of ultra-processed food consumption with cardiovascular risk factors among patients with type-2 diabetes mellitus. Nutr Diabetes 2024 Oct 22;14(1):89. [doi: 10.1038/s41387-024-00337-8] [Medline: 39433734]

XSL•FO
RenderX

Abbreviations

CGM: continuous glucose monitoring
HbA_{1c}: hemoglobin A_{1c}
HEOR: health economics and outcomes research
IoT: Internet of Things
QoL: quality of life
SMBG: self-monitoring of blood glucose
T2DM: type 2 diabetes mellitus
TAR: time above range
TBR: time below range
TIR: time in range

Edited by I Steenstra, T Leung; submitted 28.Oct.2025; peer-reviewed by A Eisingerich, D Ahn, M Chitale; revised version received 05.May.2026; accepted 19.May.2026; published 08.Jul.2026.

Please cite as:
Saboo B, Sabharwal M, Gharia M, Aswal D, Modi P, Maheshwari T, Parikh J, Saanyal S, Agrawal J
App-Based Digital Therapeutics Integrating Continuous Glucose Monitoring for Glycemic Control in Type 2 Diabetes: Prospective Observational Cohort Study
JMIR Diabetes 2026;11:e86651
URL: https://diabetes.jmir.org/2026/1/e86651
doi:10.2196/86651

© Banshi Saboo, Mudit Sabharwal, Mihir Gharia, Devina Aswal, Parth Modi, Twinkle Maheshwari, Jaymin Parikh, Simpel Saanyal, Jaikishan Agrawal. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 8.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Quality and Guideline Adherence of Mobile Nutrition Management Apps for Diabetes: Evaluation Study

Sidra Jahanzeb, MSc; Nadia Davoody, MSc, PhD

Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18 A, Stockholm, Stockholm County, Sweden

Corresponding Author:
Nadia Davoody, MSc, PhD
Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18 A, Stockholm, Stockholm County, Sweden

Abstract

Background: Diabetes mellitus is a chronic metabolic disorder marked by elevated blood glucose levels and has emerged as a global epidemic that requires management strategies for effective glycemic control through diet. In recent years, mobile apps have emerged as valuable tools for supporting self-management in chronic diseases such as diabetes, particularly for the nutritional aspects of the disease. However, the quality, accuracy, and adherence of these apps to established dietary guidelines remain underexplored and inconsistent.

Objective: The study aims to evaluate the quality and adherence to guidelines of digital nutrition management apps for diabetes, with a focus on dietary guidelines from the American Diabetes Association (ADA), World Health Organization (WHO), European Association for the Study of Diabetes (EASD), and Diabetes Canada (DC).

Methods: A quality evaluation was performed, involving the identification of mobile apps from the Google Play Store and Apple App Store. A total of 24 apps were selected based on predefined inclusion and exclusion criteria. Apps were analyzed for their content and features using a compliance checklist derived from official dietary guidelines for diabetes, including carbohydrate tracking, meal planning, glycemic control, fiber intake, physical activity, weight management, and education about diabetes. Additionally, the Mobile App Rating Scale was used to evaluate app quality in terms of engagement, functionality, aesthetics, and information. Intraclass correlation coefficients were used to calculate interrater reliability. Pearson correlation coefficients were also calculated.

Results: Only 2 apps showed full compliance with the dietary guidelines, while most apps showed partial adherence. The Mobile App Rating Scale evaluation revealed significant variability in app quality, with mean total scores ranging from 9.46 to 17.69. This finding indicated major gaps in user engagement, functionality, educational content, and personalization. The intraclass correlation coefficient was 0.86, which indicates good interrater reliability, and the Pearson correlation coefficient was 0.88, suggesting good consistency between the authors.

Conclusions: Even with the growing availability of nutrition management apps for diabetes, many lack full compliance with dietary guidelines and show room for improvement in their content quality. Collaboration between health care professionals, developers, and patients is essential for the future development of these tools to effectively support diabetes self-management. Strengthening guideline adherence and content quality can increase the effectiveness of these digital tools in promoting self-management and improving health outcomes for individuals with diabetes.

(JMIR Diabetes 2026;11:e80890) doi:10.2196/80890

KEYWORDS

diabetes mellitus; mobile apps; dietary guidelines; self-management; MARS; mHealth apps; evaluation; Mobile App Rating Scale; mobile health

Introduction

Diabetes Mellitus

Diabetes mellitus is a chronic metabolic disorder marked by elevated blood glucose levels and has emerged as a global epidemic. According to the World Health Organization (WHO), “Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys, and nerves” [1].

It is one of the leading causes of morbidity and mortality, and it is often called the “silent killer” due to its slow onset and the absence of obvious symptoms in its early stages, which can lead to severe long-term complications if left undiagnosed and untreated, including cardiovascular disease, kidney failure,

XSL•FO
RenderX

neuropathy, and retinopathy. These complications significantly affect the quality of life of individuals and lead to increased health care costs [2,3].

Role of Diet in Diabetes Management

Diabetes management is effective through combining medical treatment, lifestyle modification, and dietary changes. Dietary management is one of the most important aspects of diabetes care and emphasizes carbohydrate counting, prioritizing foods with a low glycemic index, intake of high-fiber foods, and a balanced diet of macronutrients [4,5].

Dietary guidelines for diabetes management have been developed by the American Diabetes Association (ADA), WHO, the European Association for the Study of Diabetes (EASD), and the Canadian Diabetes Association (CDA). These guidelines are recognized worldwide by health care professionals and play an important role in effective diabetes management [6]. Despite the availability of these guidelines, many individuals still struggle with dietary adherence because of a lack of education and little to no access to professional support [7].

Although the main goal of dietary management in diabetes is blood glucose control, its benefits extend beyond glycemic management. A balanced and healthy diet can help patients with diabetes manage other risk factors, such as hypertension and high cholesterol levels, which are common among people with diabetes [8].

Digital Health Solutions and Their Challenges for Nutrition Management of Diabetes

In recent years, digital tools for health care, particularly mobile health (mHealth) apps for self-managing diseases, have taken the industry by storm. These mobile apps have many functionalities, including nutritional information, meal planning, carbohydrate counting, monitoring physical activity, and medication reminders, among others [7]. With the help of these mHealth apps, patients can easily access their nutritional data, track their progress, and receive reminder notifications about their activities, which in turn may help patients remain mindful of their eating habits and other activities and improve their ability to manage their condition [9,10].

mHealth apps may have the potential to act as a bridge between patients and health care providers by providing personalized recommendations. These apps can also be a source of educational content that can help patients understand the importance of dietary choices and their effect on blood glucose levels [11]. Moreover, some apps also include a feature of barcode scanning for food products that allows users to access nutritional information about whether the product is compatible with their dietary needs. Some apps are integrated with wearable devices (eg, fitness trackers and glucose monitors) to give real-time glucose levels and feedback that further increase the ability of the user in managing their condition [12].

The use of digital tools in diabetes management is not limited to mHealth apps. Other technologies have also come into play, such as continuous glucose monitoring systems and insulin pumps, which have also been integrated with mHealth apps [13]. However, there are concerns regarding the quality, credibility, and adherence of these mHealth apps to established dietary guidelines [14].

To support diabetes management, these apps need to have a solid background that supports and aligns with diabetic nutrition guidelines, which many apps fail to do. Without this, these apps may not be able to assist patients in managing their condition, which can lead to undesirable and potentially dangerous health outcomes [15]. Therefore, a systematic evaluation of these apps in relation to official dietary guidelines is necessary to assess their functionality and identify areas for improvement.

Aim and Objective of the Study

This study aimed to evaluate the quality and adherence to guidelines of digital nutrition management apps for diabetes, with a focus on the main dietary guidelines from the ADA, WHO, EASD, and Diabetes Canada (DC). This study compared the features of various apps, assessed whether they comply with dietary guidelines, and provided recommendations for improvement.

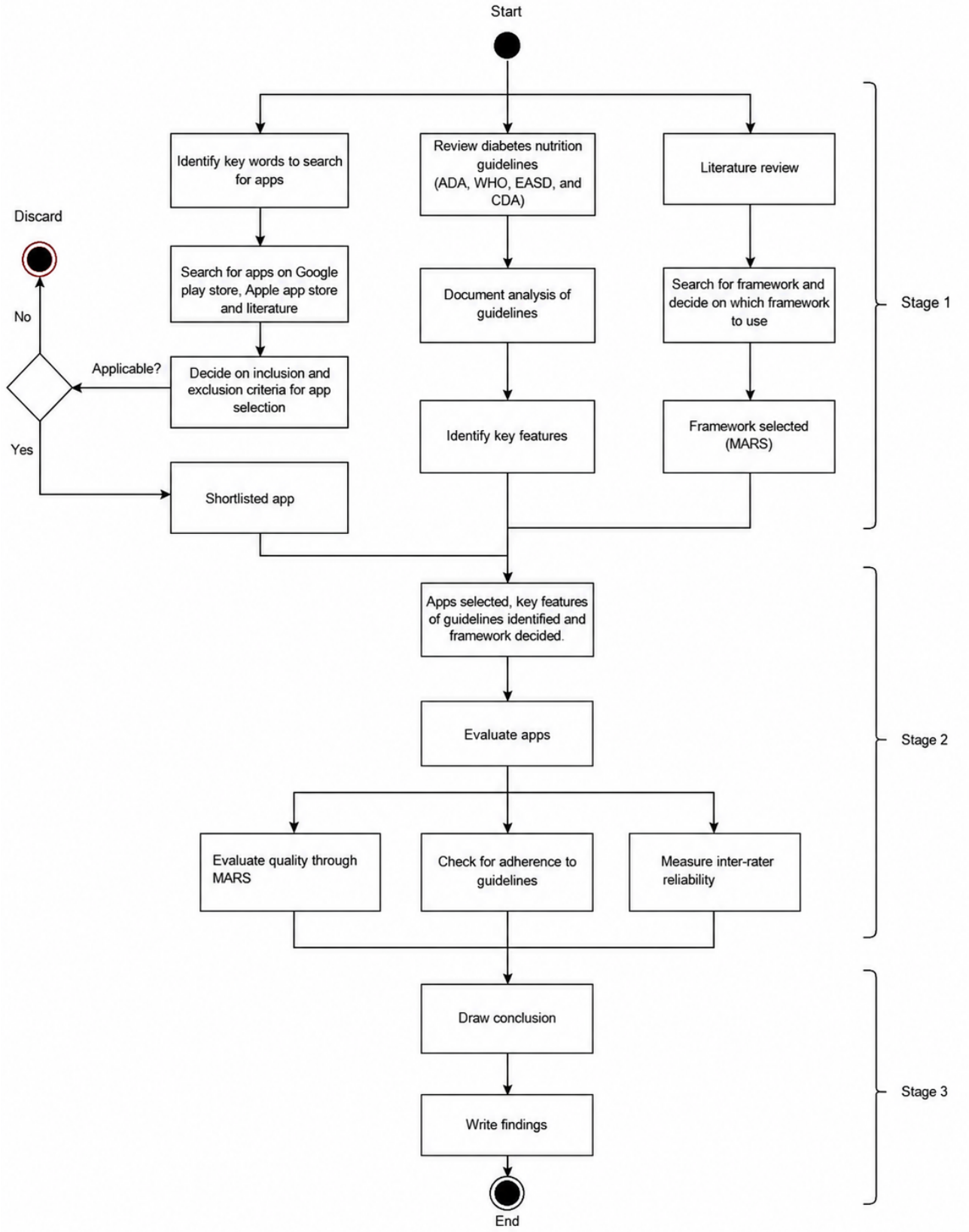
Methods

Study Design and Setting

This section provides an overview of the research methodology used in this study. It consisted of several stages where work was divided into several sections, and some of these sections took place simultaneously. A Unified Modeling Language diagram in Figure 1 shows all the phases used throughout this study.

XSL•FO
RenderX

Figure 1. Unified Modeling Language (UML) activity diagram illustrating phases of the study. ADA: American Diabetes Association; CDA: Canadian Diabetes Association; EASD: European Association for the Study of Diabetes; MARS: Mobile App Rating Scale; WHO: World Health Organization.



This evaluation study assessed the quality and guideline compliance of mHealth apps aimed at supporting nutrition management for individuals with diabetes. The research was conducted in 3 stages that involved multiple steps in each stage to achieve the desired result. The research approach was designed to ensure a thorough selection process such that no relevant app was missed, and the selected mHealth apps were functionally adequate and met the study criteria.

The study was conducted in 2025, with data collection and analysis carried out between January 17, 2025, and June 11,

XSL•FO
RenderX

2025. The primary source of app selection was the Google Play Store, Apple App Store, and existing literature, from which mHealth apps were selected based on predefined inclusion and exclusion criteria. The analysis was conducted on publicly available apps, dietary guidelines, and evaluation frameworks, with all research taking place in Stockholm, Sweden.

Data Collection

A systematic search strategy was used to identify eligible apps. The app identification process was performed in two stages: searching the Apple App Store and Google Play Store and searching the literature to identify additional apps reported in previous studies [11,12,15]. The databases used for the literature search were PubMed, Web of Science, and CINAHL. The search strategy used keywords related to diabetes, mobile apps, and nutrition management. The following search terms were used for each database: ("diabetes" OR "diabetes mellitus" OR "type 1 diabetes" OR "type 2 diabetes") AND ("mobile app" OR "mobile application" OR "mHealth" OR "smartphone app") AND ("nutrition" OR "diet" OR "meal planning" OR "carbohydrate counting" OR "food tracking" OR "diet management")

The search was performed on English-language publications only. Titles and abstracts were screened to identify relevant studies evaluating or mentioning mobile apps for diabetes nutrition management.

Similarly, the Google Play Store and Apple App Store were searched systematically for apps using predefined keywords. The search was performed in February 2025. The following keywords were used in both app stores: "diabetes," "nutrition management," "meal planning," "diabetes management," "diabetes diet," "carb counting," "diabetes meal tracking," "glucose control," "glycemic index," "diabetes recipes," "diabetes food tracker," and "diabetes nutrition app."

For apps downloaded from the Google Play Store, a Samsung S22 Ultra with Android (version 14) was used, and an iPhone 14 Pro Max with iOS (version 18.3.2) was used to download apps from the Apple App Store. The search was performed through a systematic search, with the same keywords used in the same sequence on both platforms to avoid missing any relevant apps and to ensure consistency. Apps that appeared as similar apps on the app page were also assessed. They were then reviewed through their titles, descriptions, and lists of features to gauge their relevance to the topic. After the initial search, a set of inclusion and exclusion criteria was used to filter the most relevant apps for evaluation. The inclusion and exclusion criteria are shown in Table 1.

Table . Inclusion and exclusion criteria for app selection.

Criterion	Inclusion criteria	Exclusion criteria
Focus	Apps that focus on diabetes management and nutrition.	Apps that only focus on weight loss without any diabetes or nutrition management.
Availability	Apps are available on Google Play Store and the Apple App Store.	App available on other search engines or requiring institutional access.
Language	Apps available in English.	Apps with insufficient English content are preventing evaluation.
Update	Apps that have been updated in the last 12 months.	Outdated apps and apps that crashed, failed to load, or were not functional during testing.
Category	Apps in the categories of Medical, Health and Fitness, Health Monitoring, Meal Planning, and Education.	Apps in any other category are not included in the inclusion criteria.
Payment	Free to download or apps with some in-app purchases.	Apps requiring a mandatory paid subscription to access core features needed for evaluation.
Target users	Apps intended for individuals with diabetes or general users managing diabetes.	Apps designed to be used by health care professionals only.
Others	__[a]	Apps that require a monetary subscription and cannot be accessed without it.
Others	__[a]	Apps that require external devices to work (eg, glucometer).
Others	__[a]	Duplicates from the Apple App Store and Google Play Store.

[a]Not available.

The selected apps must have been updated within the last 12 months to focus on actively maintained apps. This criterion ensured that included apps were compatible with current operating systems, received ongoing security and bug fixes, and contained up-to-date content and features. User ratings were not considered as part of the inclusion or exclusion criteria for this study due to potential bias in the ratings. User ratings are often hard to judge due to factors such as individual user experience, initial impressions, and limited knowledge about the app, rather than overall quality, content, or functionality. Moreover, user ratings or reviews may not provide an accurate picture of app quality and functionality, as most users focus more on the usability and interface design rather than the validity or accuracy of the content [16].

XSL•FO
RenderX

All identified apps were screened through a multistage process to ensure relevance and quality. Duplicates were removed, followed by exclusion of apps that did not meet predefined inclusion criteria such as focus on diabetes nutrition, language, update, and accessibility. The screening process was conducted consistently using the same search terms and criteria on both the Apple App Store and Google Play Store.

Data Analysis

The study was conducted by 2 authors (SJ and ND), both of whom have experience in eHealth design, digital self-management support systems, and evaluation of health technologies using established assessment frameworks. Before conducting the app evaluations, the authors reviewed the methodological literature on the use of the Mobile App Rating Scale (MARS) across various health domains and familiarized themselves thoroughly with its structure, scoring procedures, and appropriate use within app evaluation studies. SJ and ND collaboratively developed the study concept and design. SJ was responsible for the initial data collection, which included searching the literature and app stores and compiling the list of eligible apps. Both authors collaboratively developed a compliance checklist based on dietary guidelines to evaluate features within each app. Both authors then evaluated the included apps for compliance with dietary guidelines. In addition, each author independently assessed app quality using MARS. Interrater reliability was statistically calculated to assess the agreement between the 2 authors, and the results were subsequently reviewed and refined by both authors. The data analysis in this study was performed in 2 parts. The first part involved a review of dietary guidelines from the ADA, WHO, EASD, and DC to identify key dietary recommendations for diabetes management [1,9,17,18]. Although these guidelines differ slightly in how dietary recommendations are presented, the main elements for diabetes management remain the same. These organizations are internationally recognized and provide a gold standard in nutrition, lifestyle, and blood glucose management for diabetes treatment. These guidelines were screened for common nutrition-related recommendations for diabetes, including carbohydrate monitoring, glycemic index guidance, meal planning support, portion control, fiber intake, regular physical activity, and education to support self-management. Both authors independently extracted nutrition-related recommendations from each guideline. Extracted recommendations were then compared across guidelines, and overlapping or similar recommendations were grouped together. Through discussion, these grouped recommendations were translated into checklist items representing dietary features based on the guidelines. The content of mHealth apps was evaluated to determine whether each app complies with the guidelines. Each app was systematically assessed based on whether it included features such as carbohydrate tracking, glycemic index monitoring, meal planning, and other key components of diabetes management, as specified in the guidelines. The final checklist included the following items: carbohydrate control, meal planning, glycemic control, fiber intake, physical activity and exercise, personalized nutrition plans, and education about the disease. The apps were individually evaluated, and their adherence to each guideline was assessed by examining whether they included features corresponding to the features on the compliance checklist and whether these features were implemented appropriately. Apps that met these criteria were classified as compliant, while those lacking the necessary features were labeled as noncompliant.

The second part included evaluating the quality of the selected apps using MARS, which is a widely used tool developed by Stoyan et al [19]. It was chosen due to its defined evaluation framework, which assesses mobile apps in a structured way across different categories and is essential for understanding both content quality and user experience. A 5-point Likert scale is used, where 1 indicates inadequate and 5 indicates excellent. The scale has a total of 29 questions covering app quality, subjective quality, and app-specific ratings. A total of 23 questions were used for the analysis of engagement, functionality, aesthetics, information, and subjective quality, while 6 customizable questions were used to assess app-specific quality. Only the app's overall quality section was used for this study, which included four categories: engagement, functionality, aesthetics, and information. The selected 4 categories are provided in Textbox 1. App subjective quality was excluded because it requires user feedback, which is outside the scope of this study, and app-specific questions were also excluded. The MARS ratings were determined exclusively based on features available for free within the apps, without considering paid or premium content. Each app was individually assessed by both authors, and the average score across the 4 categories was calculated, with scores ranging from 1 to 5. Therefore, the total score was calculated by dividing the total score across all 4 categories by 4. App developers, health care professionals, and researchers are the primary users of this scale [19], and it has been consistently used in previous studies [20] to evaluate mHealth apps either alone or in combination with other tools, which supports its reliability for this type of analysis.

XSL•FO
RenderX

Textbox 1. App quality rating aspects assessed using the Mobile Application Rating Scale (MARS).

Engagement:

- Entertainment
- Interest
- Customization
- Interactivity
- Target group

Functionality:

- Performance
- Ease of use
- Navigation
- Gestural design

Aesthetics:

- Layout
- Graphics
- Visual appeal

Information quality:

- Accuracy of app description
- Goals
- Quality of information of app content
- Quality of information regarding the extent of coverage
- Visual information
- Credibility
- Evidence-based

To evaluate interrater reliability for total MARS scores, a 2-way mixed-effects ANOVA was conducted. This is an efficient statistical method available that partitions total variability in the data into separate components to determine how much variation is explained by different sources [21]. The intraclass correlation coefficient (ICC [3,1]) was derived from ANOVA mean squares and estimated absolute agreement between fixed raters. Pearson correlation between the total scores of both authors was also calculated.

The underlying ANOVA model used was as follows:

$$Yij=\mu+\alpha i+\beta j+\varepsilon ij$$

where Y_{ij} is the MARS score for app *i* by rater *j*, μ is the overall mean, α_i is the random effect of app *i*, β_j is the fixed effect of rater *j*, and ε_{ij} is the residual error.

The ICC was calculated using the following formula for a 2-way mixed-effects absolute-agreement model with single measures:

$$ICC(3,1)=MSR-MSEMSR+(k-1)MSE$$

where MRS is the mean square for apps, MSE is the mean square error, and *k* is the number of evaluators, which were 2 in this study. Values below 0.5 indicate poor reliability, 0.5 - 0.75 indicate moderate reliability, 0.75 - 0.9 indicate good reliability, and values above 0.9 indicate excellent reliability.

Ethical Considerations

The study adhered to ethical guidelines to ensure that the research was conducted responsibly. Direct human participation and personal data collection were not required because the study focused on analyzing mHealth apps for diabetes management. All selected apps were freely available on the Google Play Store and Apple App Store, and the evaluation was based on publicly available content such as app descriptions, features, and functionalities. As no personal or sensitive data from users were involved, data privacy concerns were minimal. Furthermore, the study used established dietary guidelines that were publicly available from ADA, WHO, EASD, and DC. The study clearly outlined the data sources and provided a transparent description of the methods and procedures to ensure that the research findings can be reliably reproduced. In terms of data management, all collected information was securely stored on password-protected devices, and no personal or identifiable information was recorded at any stage of this research.

XSL•FO
RenderX

Results

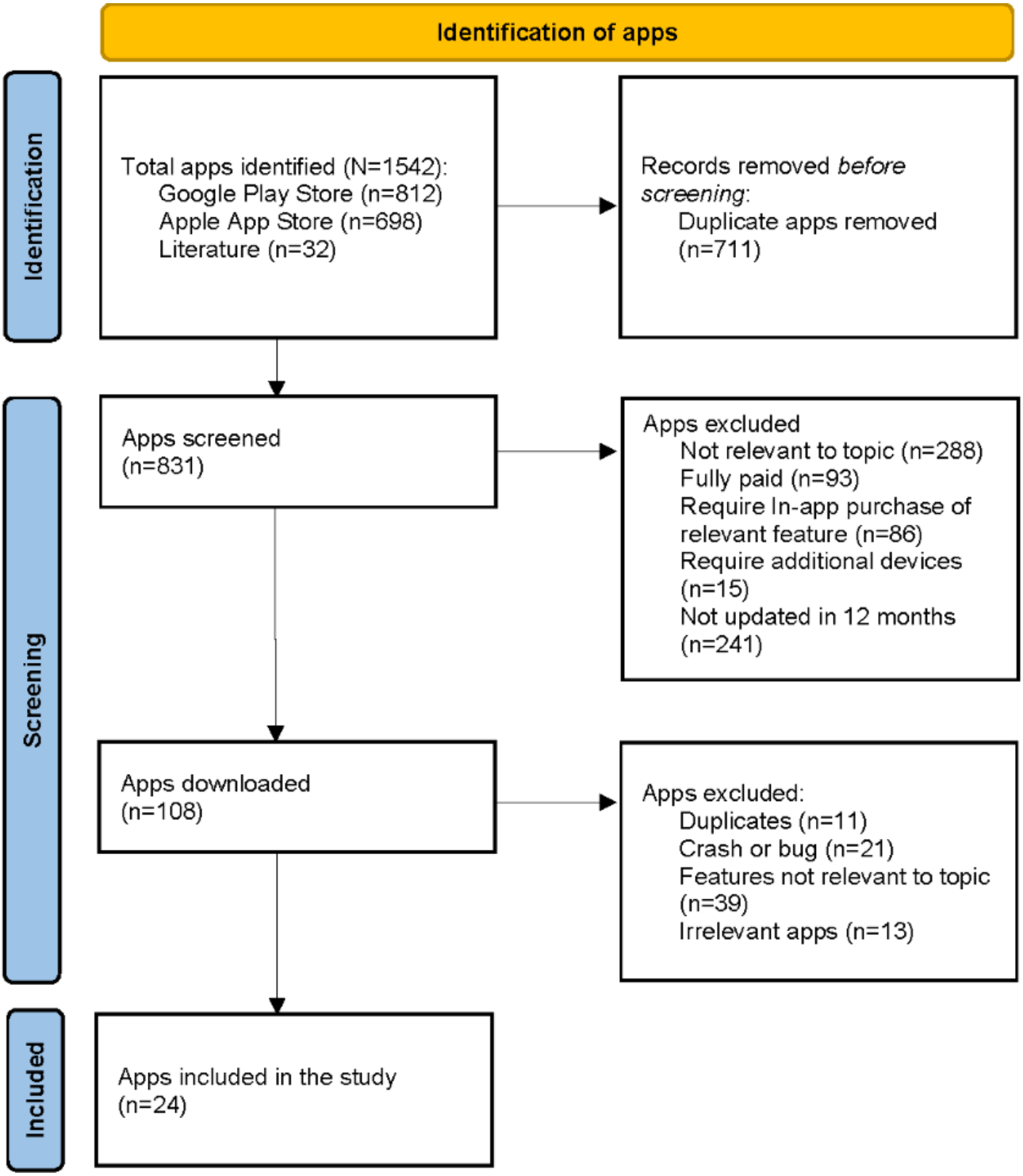
Overview

This section presents the findings of the evaluation of 24 selected mHealth apps for diabetes nutrition management. The following subsections include results for app screening, description of the apps, evaluation of compliance with dietary guidelines, and MARS evaluation.

Apps Screening

A total of 1542 mobile apps were identified, 812 from Google Play Store, 698 from Apple App Store, and 32 from previous literature. After removing 711 duplicates, 831 apps were screened for eligibility. A total of 108 apps were downloaded for further review after following the inclusion and exclusion criteria. Moreover, 84 apps were excluded due to duplication, crashes, or lack of relevant features, leaving 24 apps for the final evaluation. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for app identification and screening is summarized in Figure 2.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for app identification and screening.



XSL•FO
RenderX

Apps Features

A basic description of the selected 24 apps used in this study is shown in Table 2. The mHealth apps were selected based on the inclusion and exclusion criteria provided in Table 1. The 24 apps were downloaded from Google Play Store and Apple App Store under the categories of Health and Fitness, Medical, Health Monitoring, Meal Planning, and Food and Drink. All apps were available on both the Apple App Store and the Google Play Store. The categories reflect the main focus of each app, where Health and Fitness apps focus on general wellness, lifestyle modification, diet tracking, and physical activity. Meal Planning apps provide structured meal plans, dietary recommendations, and nutrition guidance. Medical apps are used for disease management and include clinical information, education, and treatment-related support. Health Monitoring apps focus on tracking health parameters such as blood glucose, dietary intake, and other self-management indicators. The Food and Drink category focuses on apps that mainly provide food databases, recipe collections, nutrition information, and meal-tracking features.

XSL•FO
RenderX

Table . Apps descriptions.

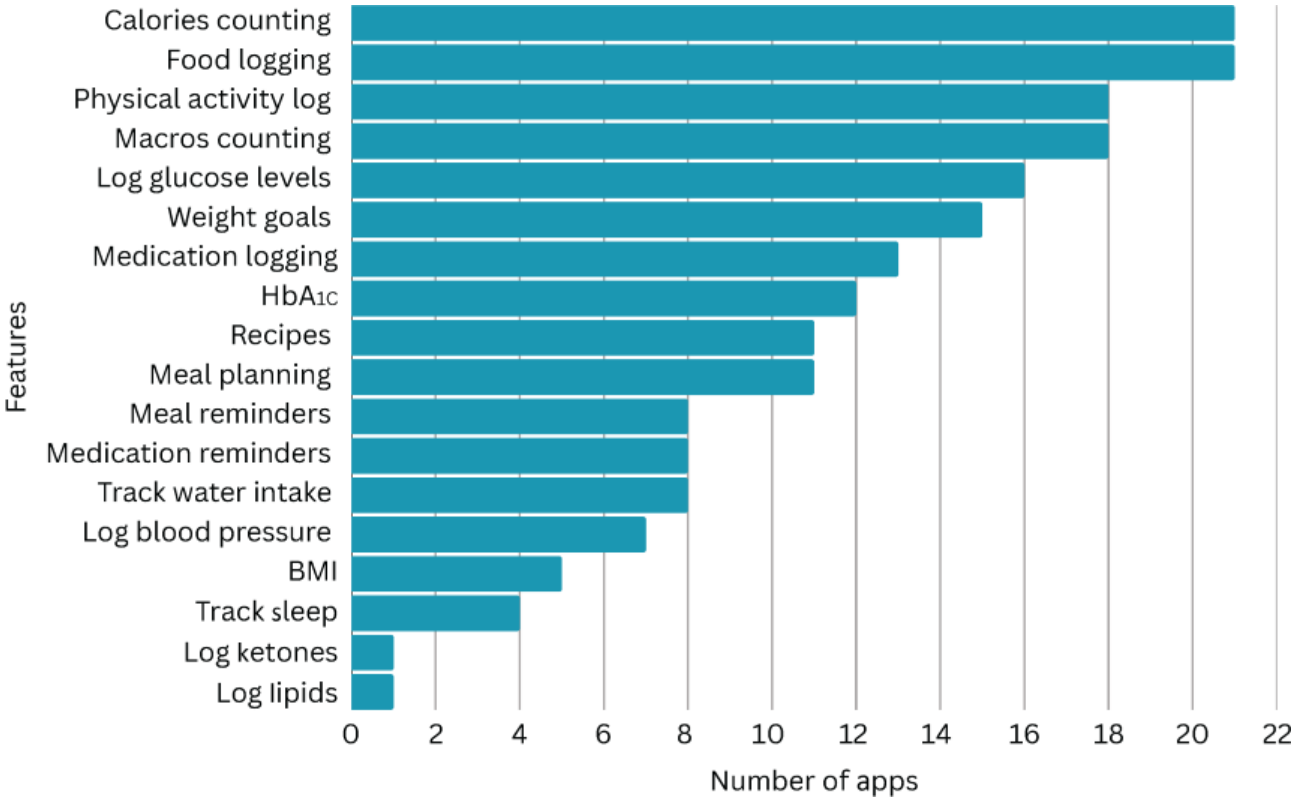
App name	App store category	Developer name	Last update	Cost
Carb Manager	Health, Fitness, and Meal Planning	Wombat Apps LLC	Feb 2025	Free but offers in-app purchases
MyDiabetes: Meal, Carb Tracker	Health, Fitness, and Medical and Health Monitoring	GO Health Solutions	Mar 2025	Free but offers in-app purchases
Yazio	Health and Fitness	YAZIO	Apr 2025	Free but offers in-app purchases
Glycemic Index. Diabetes Diary	Medical	App Holdings	Dec 2024	Free but offers in-app purchases
Health2Sync–Glucose Tracker	Medical and Health Monitoring	H2 Inc	Apr 2025	Free but offers in-app purchases
MyNetDiary	Health and Fitness	MyNetDiary.com	Apr 2025	Free but offers in-app purchases
mySugr	Medical and Health Monitoring	mySugr GmbH	Apr 2025	Free but offers in-app purchases
Glycemic Index & Load App	Health and Fitness	RFIT Development	Mar 2025	Free but offers in-app purchases
Glooko	Health, Fitness, and Medical and Health Monitoring	Glooko Inc	Mar 2025	Free
Carbs & Cals: Diet & Diabetes	Health and Fitness	Chello Publishing Limited	Apr 2025	Free but offers in-app purchases
Center Health	Medical	Center Health Inc	Apr 2025	Free
BeatO	Health, Fitness, and Medical	Health Arx Technologies Private Limited	Feb 2025	Free
DiabTrend	Health, Fitness, and Medical and Health Monitoring	DiabTrend AI Analytics Inc	Jan 2025	Free but offers in-app purchases
Diabetic Diary–Glucose Tracker	Health, Fitness, and Medical	mEL Studio	Feb 2025	Free but offers in-app purchases
Glucose Buddy	Medical and Health Monitoring	Azumio Inc	Jul 2024	Free but offers in-app purchases
Diabetes Recipes Diabetic Diet	Health and Fitness	Edutainment Ventures	Apr 2025	Free but offers in-app purchases
Fooducate	Health and Fitness	Maple Media	Nov 2024	Free but offers in-app purchases
Diabetes:M	Medical	Sirma Medical Systems	Feb 2025	Free but offers in-app purchases
Diabetes Tracker	Medical	Axel Stein	Mar 2025	Free but offers in-app purchases
SocialDiabetes	Medical and Health Monitoring	SocialDiabetes	Apr 2025	Free but offers in-app purchases
Gluroo: Diabetes Log Tracker	Health, Fitness, and Medical and Health Monitoring	Gluroo Imaginations Inc	Apr 2025	Free
nBuddy Diabetes	Health and Fitness	HeartVoice	May 2025	Free but offers in-app purchases
Blood Sugar Diary for Diabetes	Medical and Health Monitoring	MedM Inc	Apr 2025	Free but offers in-app purchases
Diabetes App–Diabetic Diet	Health, Fitness, and Food and Drink	Aquila Soft	Aug 2024	Free but offers in-app purchases

A detailed overview of the features of all the apps is provided in Multimedia Appendix 1. A bar chart illustrating various features of mHealth apps designed for nutritional management of diabetes, showing the tools users can use to monitor and manage their condition, is provided in Figure 3. The most common features were calorie counting and food logging, each present in 21 of the 24 apps, indicating that most apps prioritized food tracking and calorie counting for diabetes management.

XSL•FO
RenderX

Macros counting and physical activity logging were present in 18 apps. The feature to log glucose levels was found in 16 apps, while other features include weight goals (15 apps), medication logging (13 apps), and hemoglobin A_{1c} (HbA_{1c}) logging or estimation (12 apps). Less frequently included features, such as water intake tracking (8 apps), blood pressure logging (7 apps), and BMI tracking (5 apps), provided a more comprehensive approach to managing diabetes by addressing other aspects of health, such as hydration and cardiovascular function. The least common features, including ketone logging and lipid logging, were present in only 1 app each.

Figure 3. Features corresponding to the number of apps. HbA_{1c}: hemoglobin A_{1c}.



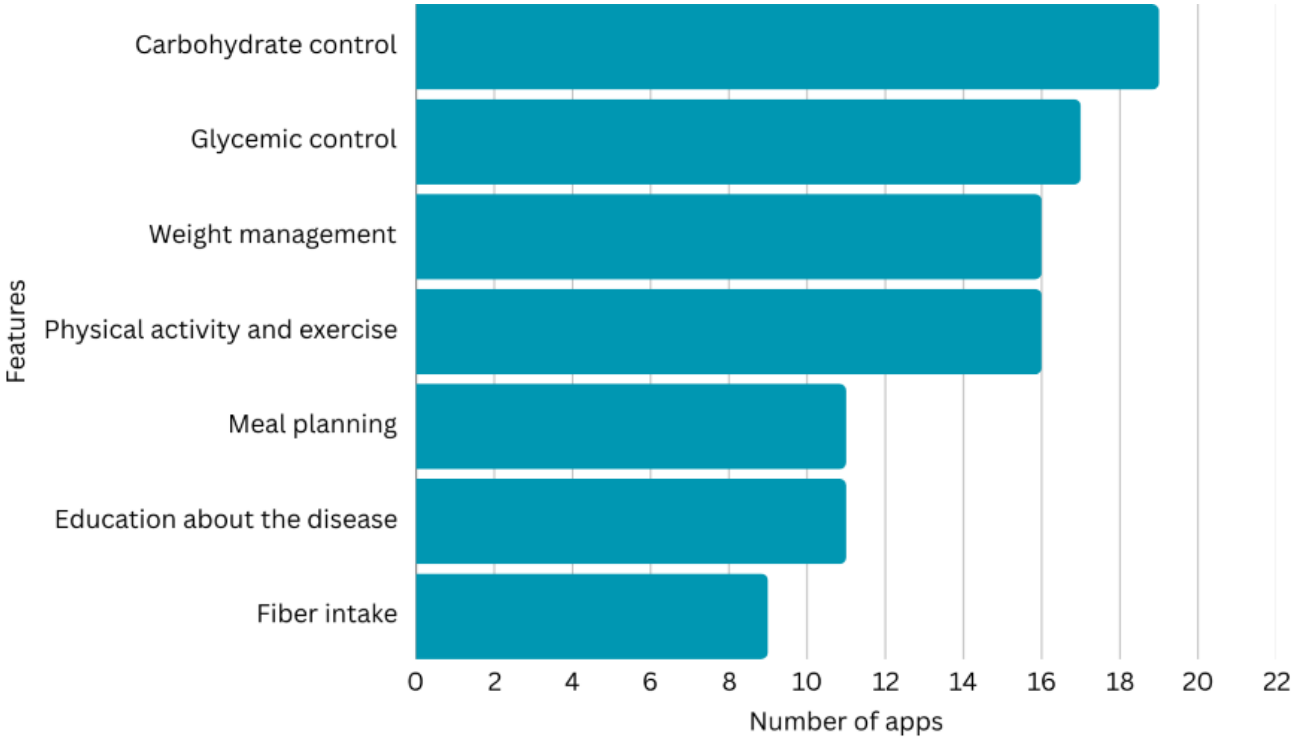
Compliance With Dietary Guidelines

The compliance of the selected 24 mHealth apps with official dietary guidelines for diabetes management was evaluated using a checklist formed by reviewing the guidelines to identify the main dietary recommendations and systematically assessing the presence or absence of these features in the apps. The analysis of the selected apps showed varying levels of compliance with essential features, with some offering proper support for diabetes management, while others displayed significant gaps.

Figure 4 illustrates the number of apps in relation to their compliance features. A detailed overview is provided in Multimedia Appendix 2, where “x” marks compliance with the dietary guidelines.

XSL•FO
RenderX

Figure 4. Compliance checklist in relation to the number of apps.



Only 2 apps, BeatO (Health Arx Technologies Private Limited) and Glycemic Index & Load App (RFIT Development), showed full compliance with dietary guidelines by including features for carbohydrate control, meal planning, glycemic control, fiber intake, physical activity, weight management, and education.

Three apps, including MyDiabetes: Meal, Carb Tracker (GO Health Solutions), Glucose Buddy (Azumio Inc), and Center Health (Center Health Inc), showed strong adherence to dietary guidelines. All these apps met 6 of the 7 checklist features that were in accordance with the dietary guidelines. For instance, MyDiabetes: Meal, Carb Tracker, and Center Health lacked only fiber intake tracking while offering all other features. Similarly, Glucose Buddy lacked only the meal-planning feature.

Five apps, including Health2Sync–Glucose Tracker (H2 Inc), Yazio (YAZIO), SocialDiabetes, DiabTrend (DiabTrend AI Analytics Inc), and nBuddy Diabetes (HeartVoice), each showed compliance with 5 of the 7 key features of the compliance checklist for diabetes management. Health2Sync–Glucose Tracker and SocialDiabetes were compliant with carbohydrate control, glycemic control, physical activity, weight management, and education about the condition, but lacked both meal planning and fiber intake tracking. The app Yazio offered strong support for carbohydrate control, meal planning, physical activity, weight management, and fiber intake, but lacked glycemic control because it did not allow logging of glucose levels and lacked education about the condition. DiabTrend and nBuddy Diabetes were both compliant with carbohydrate control, glycemic control, physical activity, fiber intake, and education about the condition, but lacked both meal planning and weight management features.

Moreover, the apps Carb Manager (Wombat Apps LLC), Gluroo (Imaginations Inc), Glooko (Glooko Inc), Carbs & Cals: Diet & Diabetes (Chello Publishing Limited), and Diabetes Recipes Diabetic Diet (Edutainment Ventures) were compliant with 4 of the 7 key features in the compliance checklist for diabetes management. All these apps provided carbohydrate control. Meal planning was present in Carb Manager, Carbs & Cals: Diet & Diabetes, and Diabetes Recipes Diabetic Diet, whereas the glycemic control feature was present in Gluroo: Diabetes Log Tracker (Gluroo Imaginations Inc) and Glooko. Furthermore, Carb Manager, Gluroo: Diabetes Log Tracker, and Carbs & Cals: Diet & Diabetes encouraged users to increase fiber intake. Weight management was offered only by Carb Manager, Glooko, and Carbs & Cals: Diet & Diabetes, while Gluroo: Diabetes Log Tracker, Glooko, and Diabetes Recipes Diabetic Diet included features to log physical activity and exercise. Finally, only Diabetes Recipes Diabetic Diet provided users with education regarding diabetes. Some apps showed partial compliance. For example, the app mySugr (mySugr GmbH) included only carbohydrate control, glycemic control, and physical activity. Similarly, Diabetes Tracker (Axel Stein), Diabetic Diary–Glucose Tracker (mEL Studio), and Glycemic Index. Diabetes Diary (App Holdings) offered features for carbohydrate control through macronutrient counting, glycemic control through tracking and logging blood glucose levels, and weight management through calorie counting and weight-goal assignment.

Additionally, MyNetDiary (MyNetDiary) and Fooducate (Maple Media) offered a unique approach to diabetes management by initially asking users whether they had diabetes or another health condition, which allowed the apps to tailor plans accordingly. These apps focused on meal planning, weight management, and physical activity. Furthermore, the apps Diabetes:M - Blood Sugar Diary (Diabetes Diet) and Diabetes App–Diabetes Diet (Aquila Soft) offered only one feature: glycemic control through logging and tracking blood glucose levels.

XSL•FO
RenderX

MARS Evaluation

The MARS was used to assess the quality of the selected 24 mHealth apps for nutritional management of diabetes. The questions used to perform the MARS evaluation are provided in Multimedia Appendix 3.

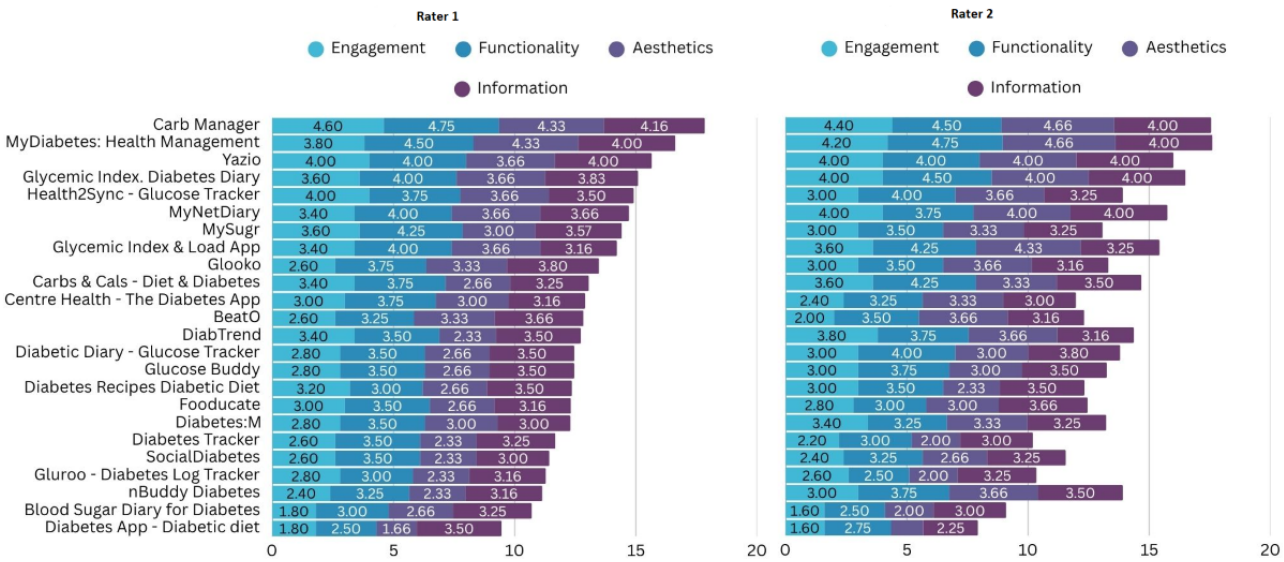
The evaluation showed significant variation in the performance of the apps. Two (8%) apps achieved high performance, with a mean total score of 17.76 for Carb Manager and 17.12 for MyDiabetes: Meal, Carb Tracker across all domains.

A larger proportion of the apps showed average performance, amounting to 67% (n=16) of the total apps, with MARS scores ranging from 12 to 16. Some apps performed reasonably well, including Yazio (15.83), Glycemic Index. Diabetes Diary (15.79), Health2Sync–Glucose Tracker (14.41), MyNetDiary (15.23), mySugr (13.75), and Glycemic Index & Load App (14.82).

Six (25%) apps showed low performance, with scores below 12. For example, the Diabetes App–Diabetic Diet scored the lowest among all apps, with a score of 9.46, followed by Blood Sugar Diary for Diabetes (9.9; MedM Inc), Gluroo: Diabetes Log Tracker (10.82), SocialDiabetes (11.49), and Diabetes Tracker (10.94).

Figure 5 shows a stacked bar chart presenting MARS scores from both authors for all apps. A detailed overview of MARS scores, including total and mean total scores for each app, is provided in Multimedia Appendix 4.

Figure 5. Mobile App Rating Scale (MARS) scores.



A 2-way mixed-effects ANOVA model was performed to assess interrater reliability for the total MARS scores, in which apps were treated as random effects and raters as fixed effects. Through this model, the scores from each app and rater were compiled into three sources of variance: variance between apps, variance between raters, and residual error. The ANOVA partitioned the total variability into these components and generated mean square values. The mean square for rows represented the differences in total MARS scores between the 24 apps, the mean square for columns represented differences between the 2 raters, and MSE represented random disagreement between raters for each app. The calculated mean square values were mean square for rows=18.68, mean square for columns=0.08, and MSE=1.08, and k=2 represented the number of raters. Using these values, the ICC (3,1) was calculated through the equation:

$$ICC(3,1)=MSR-MSEMSR+(K-1)MSE$$

$$ICC(3,1)=18.68-1.0818.68+(2-1)1.08=0.86$$

An ICC of 0.86 was calculated, which indicated good-to-excellent interrater reliability according to established benchmarks. A Pearson correlation coefficient was also calculated between the total scores of the 2 raters, yielding a value of 0.88, further indicating a strong positive relationship and good agreement and consistency.

Discussion

Principal Findings

A detailed evaluation of mHealth apps for nutritional management of diabetes was conducted in this study to assess their functionality and adherence to dietary guidelines provided by the WHO, ADA, EASD, and DC. The growing reliance on mHealth apps to support patients in managing chronic conditions, such as diabetes and hypertension, has raised concerns regarding the reliability of these digital tools [22]. A total of 24 apps were selected based on predefined inclusion and exclusion criteria. Only 2 apps fully adhered to the dietary guidelines, whereas the majority demonstrated only partial compliance. The MARS assessment indicated considerable variability in overall app quality, with mean total scores ranging from 9.46 to 17.69. These findings highlight substantial shortcomings in user engagement, functionality, educational content, and personalization. The calculated ICC of 0.86 reflected good interrater reliability, and the Pearson correlation

XSL•FO
RenderX

coefficient of 0.88 further indicates strong consistency between the reviewers.

Apps Features

There was significant variation among the evaluated mHealth apps in terms of features and functionality. Most apps focused on general tracking features such as blood glucose levels, carbohydrate intake, medication adherence, and physical activity. However, many apps lacked dietary management tools, including structured meal planning, fiber intake, and personalized dietary feedback. Although these tools have the potential to support self-management, our findings indicate that many apps lack full adherence to established dietary guidelines and do not consistently deliver personalized or evidence-based feedback.

Blood glucose tracking was the most common feature, allowing users to log or sync their readings and, in some cases, visualize trends or receive insights. Apps such as mySugr and Diabetes:M (Sirma Medical Systems) provided pattern analysis based on user-entered data, which may help users adjust their diet or medication. These findings are consistent with previous studies [23], which found that diabetes apps with blood glucose tracking and data visualization are better at promoting user engagement and self-monitoring. Similarly, trend analysis and personalized insights were key features associated with high MARS scores and user satisfaction [24]. Macronutrient counting, particularly for carbohydrates, was another frequent component. Apps such as Carb Manager and Yazio offered extensive food databases and accurate meal logging, which are important for maintaining stable blood glucose levels and reducing the risk of complications such as hyperglycemia or hypoglycemia [25,26].

Furthermore, meal planning and food logging features were present in some apps to help users with suggestions for balanced meals based on their dietary preferences. However, the quality and depth of meal-planning features varied across the apps. Some showed clear nutritional information, while others lacked in this regard and provided insufficient guidance, which may negatively affect glycemic control [27]. Additionally, the cultural relevance of meal plans was limited in many apps, as most recipes or ingredients were catered toward Western dietary patterns, which may potentially reduce usefulness for diverse populations.

Reminder systems were also included in some apps, including mySugr, MyDiabetes, and Diabetes:M, to support adherence to medication or meal schedules. Push notifications and customized reminders are common features that help users adhere to schedules related to their health management [28]. However, the frequency and customization options of reminders varied across apps. Previous studies [29] have shown that the absence of reminder features can lead to low user engagement and hinder long-term adherence to the app.

Physical activity tracking was another feature that was included in most of the evaluated apps. Regular exercise is one of the most important aspects of diabetes management, as it helps improve insulin sensitivity and contributes to better blood glucose management. Regular exercise also helps in weight control, which reduces the risk of complications associated with diabetes, such as heart disease and high blood pressure. Only a few apps, such as Diabetes:M and Blood Glucose Diary for Diabetes, did not have this feature. A study showed that not all apps designed for chronic disease management include the essential components required for effective self-management, which aligns with the findings of these studies [30,31].

Additional features found in only a few apps included recipes, tracking water intake, BMI, logging blood pressure, ketones, and lipids. While these may not be essential for diabetes management, their inclusion can enhance the overall functionality of the app and support more comprehensive disease management. Overall, the presence of essential tracking tools with optional wellness features shows that current apps emphasize monitoring but often lack the integration needed for complete management of diabetes.

Compliance With Dietary Guidelines

Maintaining blood glucose levels is important to prevent long-term complications in diabetes [32], as emphasized by major organizations including the ADA, WHO, EASD, and DC [1,9,17,18]. These guidelines consistently highlight carbohydrate control, fiber intake, glycemic monitoring, meal planning, physical activity, and patient education as the main components of diabetes management. Findings from this study showed significant variation between evaluated apps in the implementation of these recommendations, with most providing only partial adherence.

Only 2 apps, BeatO and Glycemic Index & Load App, fully met all the criteria on the compliance checklist. Several other apps, such as MyDiabetes and Glucose Buddy, provided strong carbohydrate logging but lacked features for fiber intake or detailed meal planning. Blood glucose monitoring was more commonly included, with many apps enabling users to log or sync blood glucose data, while some apps omitted this critical function. This finding is consistent with previous studies [31,33], which have shown gaps in carbohydrate and glycemic tracking that reduce the effectiveness of diabetes management apps. Inadequate carbohydrate tracking may lead to poor blood glucose management and increase the risk of complications such as diabetic neuropathy and cardiovascular disease [3]. The ability to track and analyze blood glucose levels over time improves diabetes management by providing users with information that enables them to make informed decisions about their health [33]. However, some apps, such as Yazio, Carbs Manager, and MyNetDiary, lacked a glucose logging feature, which may limit users' ability to adjust their diet, medication, or exercise for effective glycemic control.

Meal planning and fiber tracking were inconsistently addressed in the evaluated apps. Carb Manager, Yazio, Glycemic Index & Load App, and MyDiabetes enabled customizable meal planning tailored to users' medical conditions, health goals, and dietary preferences, in alignment with established dietary guidelines, while others showed limited content. Only a few apps included features that promote a fiber-rich diet, which has been shown to enhance glycemic control, reduce HbA_{1c} levels, and lower the risk of diabetes-related complications [34,35]. This result shows that the majority of apps related to diabetes

XSL•FO
RenderX

management failed to incorporate core dietary recommendations from international diabetes management guidelines [36].

Most apps included features for physical activity and weight management, consistent with guideline recommendations emphasizing exercise and weight control for effective glycemic control and cardiovascular health [37,38]. However, many apps lacked these features, limiting their effectiveness in supporting users with this critical aspect of diabetes care.

Patient education, which is another important component of diabetes care, was present in fewer than half of the evaluated apps. Educational content was delivered through brief text, visuals, or links and covered basic topics such as carbohydrate counting and physical activity. Apps without educational content may limit user understanding of how to manage their condition effectively [37], while the inclusion of educational content has been associated with improved diabetes self-management and reduced risk of complications [39].

In summary, adherence to mHealth apps for diabetes management to established dietary guidelines was variable. While some apps performed well in areas such as carbohydrate control, glycemic management, meal planning, fiber intake, weight management, physical activity, and educational content, others fell short of meeting the standards set by the ADA, WHO, EASD, and DC. Similarly, a previous study [40] found that only a small percentage of diabetes apps showed full compliance with dietary guidelines through their implemented features. In this study, only 2 of the 24 apps demonstrated full adherence to the guideline compliance checklist, which highlights a persistent issue identified in both this and previous evaluations [40]. The findings emphasize the importance of incorporating features essential for diabetes management, as prescribed by dietary guidelines, to improve the effectiveness of these apps.

MARS Evaluation

The MARS evaluation revealed variability in the quality of diabetes nutrition management apps, which showed that most apps had moderate performance and only a small portion scored highly. Overall, MARS scores among the 24 evaluated apps ranged from 9.46 to 17.69 out of a possible 20. Only 2 apps scored above 17, while one-quarter of the evaluated apps scored below 12, indicating weaknesses in user engagement, interface design, and information quality. These findings suggest that although many apps provide basic functionality, only a few have the quality required to effectively support diabetes self-management.

One of the main observations from the evaluation was that engagement was one of the weakest domains, and the apps Blood Sugar Diary for Diabetes and Diabetes App–Diabetic Diet received low scores from both authors in this category, indicating that they lacked sufficient features to keep users engaged or motivated to continue using the app. The highest score in this category was achieved by Carb Manager. Engagement is essential in encouraging users to interact with the app regularly and continue using it over time, which is important for managing chronic conditions such as diabetes. Engaging apps allow users to personalize their experience by setting goals, tracking progress, accessing interesting features, and receiving feedback, which helps in long-term adherence, increases motivation, and improves user satisfaction [41]. The absence of gamification and interactive features in many apps may further reduce sustained use. These features improve adherence and increase the frequency of blood glucose monitoring in patients with diabetes, which is consistent with the findings of this study that several apps show low performance in the engagement category [27,42].

Functionality achieved the highest mean scores among the 4 MARS domains, indicating that most apps operated reliably, were easy to navigate and performed their main tasks effectively, such as tracking carbohydrates, blood glucose, meal planning, and logging. Apps with lower scores had poorly designed user interfaces, which hindered users' ability to navigate, locate features, and input data efficiently, potentially leading to early abandonment of the app. This finding highlights the importance of a clear and user-friendly interface as a critical factor in the overall effectiveness of mobile apps [43].

The aesthetics and information categories revealed weaknesses in most apps. Although some apps had aesthetically pleasing designs, many lacked visual appeal and consistency in design, which could lead to reduced user engagement. A clear and visually pleasing design is essential for mHealth apps to encourage frequent interaction. Poor aesthetic quality can impact user retention and satisfaction, as it can make the design too cluttered, outdated, or low quality. Such apps are likely to experience reduced user engagement over time, thereby diminishing their effectiveness as tools for diabetes management [44]. Apps often have low customer adherence due to poor user interface design, lack of aesthetic features, and poor graphics [45], which is consistent with the results of this study. The quality of information was also uneven; only a few apps provided comprehensive, evidence-based dietary content, while others included limited or outdated information. Apps that have limited visual information and lack links to information resources can result in abandonment of the app by the user [46]. Educational content within mHealth apps is crucial for improving self-management in chronic conditions through accessible knowledge for users, so they can make informed decisions about their health. This aligns with findings from a previous study [47], which showed that most diabetes self-management apps receive a moderate score on the MARS scale, particularly underperforming in the engagement and information domains, consistent with the results of this study.

In summary, the MARS analysis showed that while some mHealth apps performed well in terms of functionality and information quality, many struggled in the engagement and aesthetics domains. The limited number of high-scoring apps shows substantial room for improvement in app design, particularly in user engagement, improving visual design, and incorporating educational content into mHealth apps for diabetes management. The high interrater reliability result seen in this study further strengthens the study and its findings, which support the consistency of quality assessment.

Strengths and Limitations

The study used established dietary guidelines from the ADA, WHO, EASD, and DC, which provided a framework for

XSL•FO
RenderX

evaluating the apps. Moreover, the study used MARS, which allowed for a standardized and systematic assessment of each app.

The study also has several limitations. The sample size of 24 apps may not fully represent the entire population of mobile apps available for diabetes management. There are many apps that can offer different features specific to certain groups of people, such as children with type 1 diabetes mellitus or those managing gestational diabetes. However, those were not included in the study due to the scope and selection criteria, which focused on general nutrition management apps for diabetes. Moreover, guideline compliance was checked by binary evaluation, but it sometimes required interpretation because features were evaluated based on how closely they matched the criteria, which can vary slightly across the guidelines. Overlapping principles were prioritized where slight differences were present to maintain consistency. However, developers may structure or label features differently during app development, which poses a risk of misclassification.

Furthermore, another limitation is the subjectivity involved in scoring with the MARS tool. Several questions, particularly those in the engagement and aesthetics categories, require personal judgment, and perceptions can vary from person to person.

Future Work

Future research should focus on evaluating the long-term impact of these apps on health outcomes, specifically investigating how sustained use of these diabetes management apps influences blood glucose management, HbA_{1c} levels, and patient adherence to lifestyle changes recommended by dietary guidelines [12,14]. Moreover, there is a need to increase the personalization of mHealth apps for the management of chronic diseases. Apps should not only provide general dietary advice but should also incorporate specific recommendations based on individual user data such as age, gender, cultural preference, and health conditions [27]. Future work could focus on integrating more adaptive learning algorithms that can offer precise and personalized feedback to users, which would improve the ability of the apps to cater to diverse populations [48]. Future studies could include patient or user feedback to strengthen research and reduce individual interpretation bias. Finally, cultural sensitivity should be considered in future app design to make diabetes management apps accessible and relevant to individuals from various cultural and ethnic backgrounds [6]. These apps can become more inclusive by incorporating localized dietary guidelines and understanding diverse cultural practices for global diabetes care.

Conclusion

The study evaluated nutrition management mHealth apps for diabetes, their adherence to dietary guidelines, and their quality using the MARS framework.

The results show significant variations in the quality and effectiveness of the evaluated mHealth apps. Most evaluated apps demonstrated partial adherence, with only 2 apps meeting all guideline criteria, including features such as carbohydrate tracking, blood glucose monitoring, meal planning, physical activity tracking, and educational information about the condition; others failed to meet the basic standards necessary for effective diabetes management.

Moreover, in the MARS evaluation, functionality scores were generally high, but engagement, aesthetics, and information quality were limited, indicating uneven app design and content. Several lower-scoring apps exhibited poor, nonengaging designs, lacked personalization, provided limited information, and featured low-quality graphics, all of which diminish their overall effectiveness and utility as tools for disease management.

The findings of the study showed gaps between dietary guidelines and their implementation in mHealth apps, which emphasizes the need for improvement in these digital tools.

In conclusion, mHealth apps showed high potential in supporting diabetes self-management, but there is still considerable room for improvement. Developers must focus on creating apps that are not only functional but also engaging, user-friendly, high-quality, and in accordance with established dietary guidelines for diabetes. By addressing these gaps, mHealth apps can become more effective tools for diabetes management by improving patient outcomes and increasing the quality of life for individuals with this chronic condition.

Acknowledgments

ChatGPT-3.5 (OpenAI) was used solely to assist with language refinement in this paper. All content, interpretations, and conclusions are the sole responsibility of the authors.

Funding

No external financial support or grants were received for this work.

Data Availability

All data relevant to this study are included in the paper and its multimedia appendices.

Authors' Contributions

SJ and ND collaboratively developed the study concept and design. SJ was responsible for data collection and conducted the preliminary analysis, which was subsequently reviewed and refined in consultation with ND. The scoring process was jointly

XSL•FO
RenderX

discussed and finalized through consensus between both authors. Both authors contributed to paper revision and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed overview of apps features.
[PNG File, 56 KB - diabetes_v11i1e80890_app1.png]

Multimedia Appendix 2
Detailed overview of compliance checklist.
[PNG File, 69 KB - diabetes_v11i1e80890_app2.png]

Multimedia Appendix 3
Questions used for the Mobile App Rating Scale (MARS) evaluation.
[PDF File, 131 KB - diabetes_v11i1e80890_app3.pdf]

Multimedia Appendix 4
Mobile App Rating Scale (MARS) evaluation scores.
[PNG File, 96 KB - diabetes_v11i1e80890_app4.png]

References

1. Diabetes. World Health Organization (WHO). URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2025-02-25]
2. IDF diabetes atlas. International Diabetes Federation. 2017. URL: https://diabetesatlas.org/resources/previous-editions/?utm_source=chatgpt.com [accessed 2026-05-23]
3. Todkar SS. Diabetes mellitus the ‘Silent Killer’ of mankind: an overview on the eve of upcoming World Health Day!. J Med Allied Sci 2016;6(1):39. [doi: 10.5455/jmas.214333]
4. Kaul K, Tarr JM, Ahmad SI, Kohner EM, Chibber R. Introduction to diabetes mellitus. Adv Exp Med Biol 2024;24(5):85-95. [doi: 10.1007/s11892-024-01536-4]
5. Gillespie KM. Type 1 diabetes: pathogenesis and prevention. CMAJ 2006 Jul 18;175(2):165-170. [doi: 10.1503/cmaj.060244] [Medline: 16847277]
6. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care 2011 Jun;34(6):1249-1257. [doi: 10.2337/dc11-0442] [Medline: 21617109]
7. Franz MJ, Bantle JP, Beebe CA, et al. Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. Diabetes Care 2002 Jan;25(1):148-198. [doi: 10.2337/diacare.25.1.148] [Medline: 11772915]
8. Glovaci D, Fan W, Wong ND. Epidemiology of diabetes mellitus and cardiovascular disease. Curr Cardiol Rep 2019 Mar 4;21(4):21. [doi: 10.1007/s11886-019-1107-y] [Medline: 30828746]
9. Standards of care in diabetes. American Diabetes Association. 2023. URL: https://professional.diabetes.org/standards-of-care [accessed 2025-02-25]
10. Shrivastava SR, Shrivastava PS, Ramasamy J. Role of self-care in management of diabetes mellitus. J Diabetes Metab Disord 2013;12(1):14. [doi: 10.1186/2251-6581-12-14]
11. Gong E, Zhang Z, Jin X, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 7;8(4):e14836. [doi: 10.2196/14836] [Medline: 32255432]
12. Corsica JA, Kelly MC, Bradley LE, et al. Mobile apps for diabetes self-management: an updated review of app features and effectiveness. J Behav Med 2025 Feb;48(1):137-148. [doi: 10.1007/s10865-024-00525-y] [Medline: 39489882]
13. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
14. Saha T, Del Caño R, Mahato K, et al. Wearable electrochemical glucose sensors in diabetes management: a comprehensive review. Chem Rev 2023 Jun 28;123(12):7854-7889. [doi: 10.1021/acs.chemrev.3c00078] [Medline: 37253224]
15. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65. [doi: 10.2196/jmir.1874] [Medline: 21979293]

XSL•FO
RenderX

16. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: a cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663. [doi: 10.2196/15663] [Medline: 32012107]
17. Aas AM, Axelsen M, Churuangsuk C, et al. Evidence-based European recommendations for the dietary management of diabetes. Diabetologia 2023 Jun;66(6):965-985. [doi: 10.1007/s00125-023-05894-8]
18. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 clinical practice guidelines for the prevention and management of diabetes in Canada. Can J Diabetes 2018;42(1):S1-S325. [doi: 10.1016/S1499-2671(17)31026-2]
19. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
20. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
22. Ho DKN, Chiu WC, Kao JW, et al. Reliability issues of mobile nutrition apps for cardiovascular disease prevention: comparative study. JMIR Mhealth Uhealth 2024 Sep 4;12:e54509. [doi: 10.2196/54509] [Medline: 39233588]
23. Adu MD, Malabu UH, Malau-Aduli AEO, Drovandi A, Malau-Aduli BS. Efficacy and acceptability of My Care Hub mobile app to support self-management in Australians with type 1 or type 2 diabetes. Int J Environ Res Public Health 2020 Apr 9;17(7):2573. [doi: 10.3390/ijerph17072573] [Medline: 32283659]
24. Lee JL, Kim Y. Evaluation of mobile applications for patients with diabetes mellitus: a scoping review. Healthcare (Basel) 2024 Jan 31;12(3):368. [doi: 10.3390/healthcare12030368] [Medline: 38338253]
25. Lopes Souto D, Lopes Rosado E. Use of carb counting in the dietary treatment of diabetes mellitus. Nutr Hosp 2010;25(1):18-25. [Medline: 20204251]
26. Kawamura T. The importance of carbohydrate counting in the treatment of children with diabetes. Pediatr Diabetes 2007 Oct;8 Suppl 6(S6):57-62. [doi: 10.1111/j.1399-5448.2007.00287.x] [Medline: 17727386]
27. Amiri M, Li J, Hasan W. Personalized flexible meal planning for individuals with diet-related health concerns: system design and feasibility validation study. JMIR Form Res 2023 Aug 3;7:e46434. [doi: 10.2196/46434] [Medline: 37535413]
28. Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J. A smartphone app to improve medication adherence in patients with type 2 diabetes in Asia: feasibility randomized controlled trial. JMIR Mhealth Uhealth 2019;8(4):e14914. [doi: 10.2196/18411]
29. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
30. Kanaley JA, Colberg SR, Corcoran MH, et al. Exercise/physical activity in individuals with type 2 diabetes: a consensus statement from the American College of Sports Medicine. Med Sci Sports Exerc 2022 Feb 1;54(2):353-368. [doi: 10.1249/MSS.0000000000002800] [Medline: 35029593]
31. Jiménez-Muñoz L, Gutiérrez-Rojas L, Porras-Segovia A, Courtet P, Baca-García E. Mobile applications for the management of chronic physical conditions: a systematic review. Intern Med J 2022 Jan;52(1):21-29. [doi: 10.1111/imj.15081] [Medline: 33012045]
32. Yazdanpanah S, Rabiee M, Tahriri M, et al. Evaluation of glycated albumin (GA) and GA/HbA1c ratio for diagnosis of diabetes and glycemic control: a comprehensive review. Crit Rev Clin Lab Sci 2017 Jun;54(4):219-232. [doi: 10.1080/10408363.2017.1299684] [Medline: 28393586]
33. Kim Y, Lee H, Seo JM. Integrated diabetes self-management program using smartphone application: a randomized controlled trial. West J Nurs Res 2022 Apr;44(4):383-394. [doi: 10.1177/0193945921994912] [Medline: 33655794]
34. Reynolds AN, Akerman AP, Mann J. Dietary fibre and whole grains in diabetes management: systematic review and meta-analyses. PLoS Med 2020 Mar;17(3):e1003053. [doi: 10.1371/journal.pmed.1003053] [Medline: 32142510]
35. Saboo B, Misra A, Kalra S, et al. Role and importance of high fiber in diabetes management in India. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2022 May;16(5):102480. [doi: 10.1016/j.dsx.2022.102480]
36. Breland JY, Yeh VM, Yu J. Adherence to evidence-based guidelines among diabetes self-management apps. Transl Behav Med 2013 Sep;3(3):277-286. [doi: 10.1007/s13142-013-0205-4] [Medline: 24073179]
37. Engin A. The definition and prevalence of obesity and metabolic syndrome. Adv Exp Med Biol 2017;960:1-17. [doi: 10.1007/978-3-319-48382-5_1] [Medline: 28585193]
38. Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015 Jul;30(7):529-542. [doi: 10.1007/s10654-015-0056-z] [Medline: 26092138]
39. Coppola A, Sasso L, Bagnasco A, Giustina A, Gazzaruso C. The role of patient education in the prevention and management of type 2 diabetes: an overview. Endocrine 2016 Jul;53(1):18-27. [doi: 10.1007/s12020-015-0775-7] [Medline: 26494579]
40. Darby A, Strum MW, Holmes E, Gatwood J. A review of nutritional tracking mobile applications for diabetes patient use. Diabetes Technol Ther 2016 Mar;18(3):200-212. [doi: 10.1089/dia.2015.0299] [Medline: 26950679]

XSL•FO
RenderX

41. Garett R, Chiu J, Zhang L, Young SD. A literature review: website design and user engagement. Online J Commun Media Technol 2016 Jul;6(3):1-14. [Medline: 27499833]
42. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 8;14(3):e70. [doi: 10.2196/jmir.2058] [Medline: 22564332]
43. Geirhos A, Stephan M, Wehrle M, et al. Standardized evaluation of the quality and persuasiveness of mobile health applications for diabetes management. Sci Rep 2022 Mar 7;12(1):3639. [doi: 10.1038/s41598-022-07544-2] [Medline: 35256661]
44. Perrig SAC, Ueffing D, Opwis K, Brühlmann F. Smartphone app aesthetics influence users' experience and performance. Front Psychol 2023;14:1113842. [doi: 10.3389/fpsyg.2023.1113842] [Medline: 37388645]
45. Liashiedzi GK, Eto FE, Atinga RA, Abor PA. Determinants of mobile health (M-Health) application adoption, usage and discontinuity among corporate workers diagnosed with hypertension and diabetes. J Health Organ Manag 2024 Mar 5. [doi: 10.1108/JHOM-11-2022-0335] [Medline: 38437500]
46. Davis DW, Logsdon MC, Vogt K, et al. Parent education is changing: a review of smartphone apps. MCN Am J Matern Child Nurs 2017;42(5):248-256. [doi: 10.1097/NMC.0000000000000353] [Medline: 28639997]
47. Krishnamurthy K, Selvaraj N, Gupta P, et al. Benefits of gamification in medical education. Clin Anat 2022 Sep;35(6):795-807. [doi: 10.1002/ca.23916] [Medline: 35637557]
48. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]

Abbreviations

ADA: American Diabetes Association
CDA: Canadian Diabetes Association
DC: Diabetes Canada
EASD: European Association for the Study of Diabetes
HbA_{1c}: hemoglobin A_{1c}
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
mHealth: mobile health
MSE: mean square error
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by S Li; submitted 18.Jul.2025; peer-reviewed by AS Brazeau, O Ajileye; revised version received 30.Mar.2026; accepted 02.May.2026; published 01.Jun.2026.

Please cite as:
Jahanzeb S, Davoody N
Quality and Guideline Adherence of Mobile Nutrition Management Apps for Diabetes: Evaluation Study
JMIR Diabetes 2026;11:e80890
URL: https://diabetes.jmir.org/2026/1/e80890
doi:10.2196/80890

© Sidra Jahanzeb, Nadia Davoody. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 1.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring User Experiences With the Lift App for Emotional Well-Being Among Youth With Type 1 Diabetes: Qualitative Study

Katie M Babbott[1], PhD; Kristin Harrison Ginsberg[1], PhD; Anna L Boggiss[2], PhD; Joanna McClintock[3], PGDipClinPsych, PhD; Chloe Hudson[4], PGDipClinPsyc; Paul Hofman[5], MBChB, FRACP; Ryan Paul[3,6], MBChB, FRACP; Alana Cavadino[7], PhD; Hiran Thabrew[1], PhD, FRACP, FRANZCP; Craig Jefferies[5,8], MBChB, FRACP; Martin de Bock[9], PhD, FRACP; Jennifer L Sherr[10], MD, PhD; Anna Serlachius[1], PhD

[1]Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand

[10]Department of Pediatrics, Yale School of Medicine, Yale University, New Haven, CT, United States

[2]Rio Tinto Children's Diabetes Centre, The Kids Research Institute Australia, Perth Children's Hospital, Perth, Australia

[3]Waikato District, Te Whatu Ora Health New Zealand, Hamilton, New Zealand

[4]Canterbury District, Te Whatu Ora Health New Zealand, Christchurch, New Zealand

[5]Liggins Institute, University of Auckland, Auckland, New Zealand

[6]School of Health Equity and Innovation, Division of Health, University of Waikato, Hamilton, New Zealand

[7]Epidemiology and Biostatistics, University of Auckland, Auckland, New Zealand

[8]Starship Child Health, Te Toka Tumai Auckland, Te Whatu Ora Health New Zealand, Auckland, New Zealand

[9]University of Otago, Christchurch, Otago, New Zealand

Corresponding Author:
Anna Serlachius, PhD
Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand

Abstract

Background: Individuals living with type 1 diabetes (T1D) are at an increased risk of experiencing psychological distress; however, there remains a scarcity of scalable and widely accessible support services, particularly for adolescents and young adults. To address this gap, digital mental health interventions are becoming an increasingly important area of innovation in diabetes care.

Objective: This study aimed to explore qualitative feedback regarding the “Lift: Thriving with Diabetes” (*Lift*) well-being app, designed to support emotional well-being among adolescents and young adults with T1D, which was recently tested in a 12-week feasibility trial conducted in New Zealand and the United States.

Methods: Of the 59 adolescents and young adults and 22 support people who participated in the main *Lift* feasibility trial, 13 agreed to participate in this secondary qualitative study. Participants attended a virtual focus group or 1-on-1 interview to discuss their experiences using the app and to explore their perspectives on the app’s engagement, functionality, and perceived impact on well-being and diabetes-related coping. Transcribed audio recordings were analyzed using directed content analysis, guided by the Mobile Application Rating Scale end-user framework (with topics of engagement, functionality, aesthetics, and information quality) and interpreted from a realist theoretical position.

Results: In total, 9 adolescents and young adults (mean age 21.5, SD 2.06 years; n=5, 56% men) and 4 support people (2 fathers, 1 friend, and 1 partner; mean age 31.3, SD 18.92 years; n=2, 50% men) completed interviews. Overall, participants viewed *Lift* as engaging, easy to use, and emotionally impactful. The most positive feedback focused on the app’s interactive features, particularly a well-being tree that “grew” with increased engagement, and its “calming” visual aesthetics. Users also reported meaningful emotional or behavioral impact, particularly in promoting connection, self-awareness, and practical coping strategies in living with T1D. However, user feedback also highlighted areas for improvement, including the need for improved content pacing, personalization, connection with other digital health tools, and greater gamification to sustain long-term engagement. Participants consistently expressed a desire for content tailored to their age, role (eg, support person vs young person), and personal preferences (eg, voice, pace, tone, and interactivity).

Conclusions: Findings underscore the potential of user-driven, emotionally intelligent digital tools to enhance well-being and connection for young people with T1D, as well as their support people. These insights can inform the refinement of *Lift* and the development of broader digital health interventions aimed at promoting well-being and fostering meaningful, sustained impact.

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e79896) doi:10.2196/79896

KEYWORDS

digital health; well-being app; type 1 diabetes; adolescents and young adults; diabetes distress; qualitative study

Introduction

Background

Individuals living with type 1 diabetes (T1D) are at increased risk for psychological distress [1,2], with notably higher rates of psychological and psychiatric disorders than those in the general population [3-5]. These rates are particularly pronounced among adolescents and young adults, whose unique developmental stage involves navigating tasks such as increasing autonomy and peer integration, which often conflict with the structured and demanding nature of diabetes management [6]. Such psychosocial challenges both disrupt and are exacerbated by the self-management requirements of T1D, creating a burdensome and often isolating cycle of distress [7,8].

Despite the high prevalence of psychosocial distress in this population, there remains a scarcity of targeted mental health services for adolescents and young adults with T1D. Existing supports are often fragmented, poorly integrated into routine care, or inaccessible due to geographical, financial, or systemic barriers [9,10]. Moreover, the emotional toll of T1D extends beyond those diagnosed; caregivers, siblings, and other loved ones often experience significant psychological strain, further impacting family functioning and the support environment for the young person [11-13]. This highlights a pressing need for interventions that address not only the clinical management of T1D but also the broader psychosocial challenges faced by adolescents and young adults and their families.

Digital Health Interventions in Diabetes Care

Digital health interventions have been widely used as an avenue of health promotion and behavior change across a variety of health conditions [14,15] and are emerging as an increasing focus of innovation in diabetes care [16]. For diabetes, there exist a handful of commercially available apps, although these are largely designed to support the achievement of glycemic targets measured by hemoglobin A_{1c} (HbA_{1c}) rather than the management of well-being more broadly [17]. Indeed, there is a gap in the digital health market for well-being apps specifically designed for individuals with diabetes.

Broadly, digital tools are well-suited to adolescents and young adults, who are often referred to as “digital natives” [18] due to their lifelong exposure to and comfort with technology. This familiarity provides a valuable opportunity to deliver engaging, scalable, and accessible interventions that align with the lifestyles and communication preferences of adolescents and young adults. In this context, digital platforms offer a compelling opportunity for delivering psychosocial support, facilitating education, and promoting self-management in a way that is both developmentally appropriate and responsive to the realities of living with T1D. In exploring methods for supporting self-management, well-being, and education, particularly among youth, digital tools are a growing solution.

In developing engaging and efficacious digital health interventions, the integration of end-user feedback is essential [19]. Such approaches allow a more granular understanding of what users want and what will keep them engaging with the interventions in a meaningful way. Maintaining user engagement, uptake, and adherence is a perennial challenge in digital health, especially among adolescents and young adults, who may quickly disengage from tools they perceive as irrelevant, unappealing, or burdensome [20].

In response to such challenges, increasing extant work has been conducted to assess digital health interventions. The Mobile Application Rating Scale seeks to assess the quality of mobile health interventions, while the uMARS (end-user version of the Mobile Application Rating Scale) framework is an end-user version of this tool [21]. By evaluating domains such as engagement, functionality, aesthetics, and information quality, the uMARS enables insight into which features are likely to support (or undermine) user interaction with an app. When used alongside co-design methods, tools such as uMARS can guide the iterative development of digital interventions, helping to ensure that they are both evidence-based and user-informed. Such approaches not only improve the likelihood of initial engagement but also enhance long-term adherence by fostering greater relevance, usability, and acceptability through direct collaboration with intended end users and ensuring that their lived experiences, needs, and preferences shape the intervention from the outset.

Rationale for This Study

“Lift: Thriving with Diabetes” (*Lift*) is a digital well-being app designed to assist young people living with T1D and the people who support them (eg, parents, friends, and partners), which was recently tested for feasibility and usability in a 12-week pilot study with 59 adolescents and young adults and 22 support people (SPs) in New Zealand (NZ) and the United States [22]. It demonstrated promising findings in terms of safety, acceptability, and engagement, with retention and completion rates greater than 80%. The pilot study also demonstrated promising within-group improvements in psychological outcomes over time. The development of *Lift* was driven by a recognized unmet need for accessible psychosocial support for adolescents and young adults within the T1D community. Few digital interventions have been specifically designed to address the well-being of adolescents and young adults with T1D, and fewer still have been developed in collaboration with the young people and their support networks whom they are intended to serve.

This study aimed to qualitatively explore user feedback of *Lift*, focusing on both individual adolescent and young adult users and adolescent and young adult–SP dyads. Specifically, this work sought to understand users’ perspectives on the app’s engagement, functionality, and perceived impact on well-being and diabetes-related coping. By capturing rich, in-depth accounts

XSL•FO
RenderX

of user experience, this study provides early insights into the acceptability and potential utility of a psychosocially oriented digital intervention for adolescents and young adults living with T1D, informing the refinement of *Lift* and similar well-being apps.

Methods

Participants

Lift was recently tested in a 12-week feasibility trial with 59 adolescents and young adults and 22 SPs in NZ and the United States [22]. In total, 13 participants (n=9, 15% adolescents and young adults and n=4, 18% SPs) from the feasibility trial agreed to participate in this secondary qualitative study.

Intervention

"Lift: Thriving with Diabetes" is an adaptation of an existing well-being app [23] that was subsequently tailored for youth living with T1D [24]. *Lift* offers an evidence-based intervention consisting of seven core modules: (1) "Feel," (2) "Relax," (3) "Be Kind to Yourself," (4) "Gratitude," (5) "Connect," (6) "Look After Your Body," and (7) "Goal Setting." Taken together, the modules aim to enhance users' overall well-being and are developed based on strategies from cognitive behavioral therapy, mindful self-compassion, psychoeducation, positive psychology, and mindfulness. The original app has shown efficacy in improving psychological outcomes in adolescents and young adults without diabetes in a randomized controlled trial [23], and further information about the content and theory underpinning *Lift* is available in a protocol paper [25].

On the basis of feedback from 34 adolescents and young adults with T1D and health care professionals working in diabetes in NZ [24], the 7 modules of the original app were modified to be diabetes-specific, lived experience videos were included, and an SP module was developed. The gamification component was also enhanced, with the badges in the original app being replaced with a "well-being tree" that grows as modules and exercises are completed. *Lift* has both NZ and US versions, with culturally congruent imagery, phrases, and voice-overs for each country.

Study Design

This qualitative study used 1-on-1 interviews and focus groups to evaluate user experiences of *Lift*, analyzed using directed content analysis. Recruitment occurred between July 2023 and December 2023. Both adolescent and young adult participants (aged 16 to 25 years) and SPs provided assent and/or informed consent and were recruited from sites within NZ and the United States in the context of their participation in a larger-scale feasibility trial of *Lift* [22]; participants were approached via email to ascertain interest in participating in this secondary qualitative study. In NZ, young people aged ≥16 years were able to provide their own informed consent. In the United States, participants aged <18 years provided assent, with informed consent obtained from a parent or caregiver.

Focus Group and Individual Interview Facilitators

Participants could choose whether they wanted to participate in a focus group or an individual interview to accommodate scheduling needs. The focus groups in this study were cofacilitated by study authors (KMB [female] and ALB [female]), and 1-on-1 interviews were facilitated independently by ALB. At the time of data collection, both study authors had prior experience in facilitating group sessions with adolescents and conducting focus groups. Both KMB and ALB hold doctorate degrees in health psychology and are registered health psychologists. Neither KMB nor ALB had any prior relationship with the participants.

Procedure

Both adolescents and young adults and caregivers who agreed to participate in this secondary study attended virtual focus groups or 1-on-1 interviews. A semistructured interview schedule (Table S2 in Multimedia Appendix 1), based on the uMARS framework [21], was developed to guide participants in discussing their experience of using the app; however, participants could discuss other diabetes-related topics outside the uMARS framework if desired. Focus groups and 1-on-1 interviews were conducted online using Zoom (Zoom Communications Inc) video conferencing; both audio and video were recorded. Transcribed audio recordings of the interviews and focus groups were analyzed using directed content analysis. Recruitment for this study was closed after data saturation was reached.

Data Analysis

Given that predefined open-ended questions from the uMARS framework were used to explore engagement with the app, directed content analysis, a qualitative method guided by existing frameworks or theories [26], was used, along with a realist theoretical position [27]. NVivo (version 15; Lumivero LLC) was used to facilitate content analysis of qualitative data, and the coding tree was housed in Microsoft Excel. Analysis was abductive, with predetermined deductive codes from the uMARS framework and inductive codes flexibly added during the coding process as needed, which is in alignment with directed content analysis methods and recommended when exploring user experience with digital health interventions [28]. Initial deductive codes were agreed upon by the research team, and 2 researchers (KB and KHG) coded the data. Following initial deductive coding, the 2 coders completed additional inductive coding and then met with the core research team to agree on final categories. Any coding disagreements were resolved via consensus between the coders, and categories were finalized through discussion until agreement was reached within the research team.

Ethical Considerations

The Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Checklist 1) was followed to report the qualitative study findings. All research methods were approved by the Health and Disability Ethics Committees in NZ (2023 EXP 13982) and the Yale Institutional Review Board (2000034956) prior to study advertisement and recruitment. Informed written consent or assent was provided by participants in advance of their participation in this study. All data were deidentified. Individuals were provided with a NZ $50 (US $28) gift voucher to compensate them for their participation.

XSL•FO
RenderX

Results

Participant Characteristics

A total of 13 participants were enrolled in this study. Of the 81 participants (n=59, 73% adolescents and young adults and n=22, 27% SPs) initially approached from the *Lift* feasibility study [22], 16 (20%) indicated interest in participating in focus groups or interviews. Subsequently, 3 (19%) participants did not respond to prompts for scheduling. Data saturation was reached after approximately 12 (92%) participants had participated in focus groups or interviews; therefore, no additional potential participants were contacted. Demographic information for the sample, stratified by adolescent and young adult (n=9, 69%) and SPs (n=4, 31%), can be found in Table 1.

Table . Demographic characteristics of the participants.

Variables	AYA[a] participants (n=9)	Support people (n=4)
Age (years), mean (SD)	21.50 (2.06)	35.25 (18.92)
Gender, n (%)		
Man	5 (55.56)	2 (50.00)
Woman	3 (33.33)	2 (50.00)
Nonbinary	1 (11.11)	0 (0)
Ethnicity, n (%)		
Chinese	1 (11.11)	0 (0)
Indian	1 (11.11)	0 (0)
Māori[b]	1 (11.11)	1 (25.00)
New Zealand European	2 (22.22)	3 (75.00)
White American	4 (44.44)	0 (0)
Relation to AYA, n (%)		
Father	N/A[c]	2 (50.00)
Partner	N/A	1 (25.00)
Friend	N/A	1 (25.00)

[a]AYA: adolescent and young adult.

[b]Indigenous people of New Zealand.

[c]N/A: not applicable.

Two 1-on-1 interviews were conducted with adolescents and young adults who could not attend the focus groups. Three focus groups were conducted: one with adolescents and young adults (n=3, 33%), one with SPs (n=3, 75%), and one combined group that included both adolescents and young adults and SPs (n=5, 38%). On average, the focus groups were 64 (SD 8.59) minutes in duration and 1-on-1 interviews were 44 (SD 1.53) minutes in duration.

User Perspectives: Engagement

The "Engagement" domain of the uMARS framework was operationalized as factors relating to entertainment, interest, customization, and interactivity and was divided into 2 subcategories: "more engaging" and "less engaging." Broadly, "Engagement" was the largest source of data from participants and was a positively skewed domain. Many participants (n=10, 77%) reported feeling positively engaged, citing features such as the well-being tree, interactive modules, journaling, and check-ins as motivating and rewarding:

> *I quite enjoyed like the check-ins...you had to take a minute to stop and reflect, and actually...recognise how you were feeling.* [SP (father), 52 years, NZ]

The visual growth of the tree, in particular, was repeatedly described as an effective incentive to return to the app. Interactive elements, such as prompts following videos, repeatable activities, and opportunities for self-reflection, were appreciated for fostering personal accountability and emotional engagement:

> *I liked how interactive it was and how you could repeat activities. It meant to me that if I wanted the comfort of doing it over, I could easily do that.* [Adolescent and young adult, 18 years, United States]

However, many participants (n=8, 62%) expressed declining engagement over time, particularly after completing all available modules:

> *It got boring rather quickly. Once all the modules are done and you figure out all the ones that work for you personally, there isn't...much else the app can really provide from that point onwards. Plus, once the tree reaches its max height, there is even less incentive to go onto the app.* [Adolescent and young adult, 17 years, NZ]

The pace of content delivery was also a barrier to engagement; several participants indicated that they wanted to speed up the

XSL•FO
RenderX

delivery of content (n=6, 46%). Although some appreciated the reflective function of check-ins, others found the repetition disengaging:

> *Having to check in every night became annoying.* [Adolescent and young adult, 22 years, United States]

Some users felt that the tone of the app, especially the narration and emotional identification tools, was overly simplistic or better suited to a younger audience (n=4, 31%). Only 2 (15%) participants noted a preference for alternative content formats (eg, written text instead of video) and requested features such as subtitles. A common thread among disengaged users was a lack of novelty and progression. Despite these criticisms, more participants reported feeling more engaged than disengaged overall.

User Perspectives: Functionality

Functionality, defined as performance, ease of use, navigation, and gestural design, was generally rated positively, with some participants (n=5, 38%) describing the app as easy to use and intuitively designed. The setup process was straightforward for nearly all users:

> *I didn't have too much trouble with it with the downloading and setting it up, it was relatively user-friendly.* [SP (father), 54 years, NZ]

Participants appreciated the dashboard layout, the ability to track progress, and the flexible structure that allowed them to complete modules in their preferred order:

> *It's easy to navigate, so you know what to do, what you've done, what's coming up. I liked that aspect of it.* [Adolescent and young adult, 22 years, United States]

However, a subset of participants reported issues with navigation (n=2, 15%) and technical performance (n=2, 15%):

> *I found it confusing to know what modules I needed to do when.* [Adolescent and young adult, 17 years, NZ]

In total, 2 (15%) users experienced app crashes or problems with audio playback. Lack of accessibility features, such as video subtitles or text alternatives, was another concern:

> *I wish it was transcribed...I didn't want everyone else to know what I'm doing.* [SP (father), 51 years, NZ]
>
> *I prefer to read rather than listen.* [Adolescent and young adult, 22 years, Unites States]

Overall, the app's functionality was viewed as a strength, particularly in terms of layout, usability, and user autonomy, although minor navigation difficulties and accessibility limitations were noted by some participants.

User Perspectives: Aesthetics

In terms of aesthetics (factors relating to layout, graphics, and visual appeal), data were divided into 2 subcategories: "positive" and "negative." Many participants (n=7, 54%) expressed positive views on the app's design, particularly its visual appeal and layout. The use of green was frequently highlighted (n=5, 38%) as contributing to a peaceful and comforting aesthetic:

> *I think the app has quite a gentle element of design to it and I think it was quite peaceful.* [Adolescent and young adult, 20 years, NZ]
>
> *It's...made to look very friendly and assuring.* [Adolescent and young adult, 19 years, NZ]

Some users also reflected that they liked the voices used within the app (n=4, 31%). Some NZ participants noted, in particular, that they valued the personalization of the app using commonplace language and voice actors with NZ accents:

> *I liked that they seemed to be Kiwi...It definitely seemed to be...in touch with New Zealand.* [Adolescent and young adult, 19 years, NZ]
>
> *I liked how they incorporated the use of te reo [indigenous language of New Zealand].* [Adolescent and young adult, 16 years, NZ]

However, some participants found these components to be negative; several participants noted that they did not find the colors appealing (n=3, 23%). Additionally, although many enjoyed the tone of the narration, others found the voices overly rehearsed (n=5, 38%):

> *Me personally I don't like the..."calming voice." It's a personal preference of mine I don't like people talking to me in that certain tone.* [Adolescent and young adult, 20 years, United States]

A small number of participants (n=2, 15%) also suggested that the visuals could more directly reflect diabetes-related themes to enhance the app's relevance to its target audience.

User Perspectives: Information

The "Information" domain of the uMARS framework, defined by content quality, clarity, and perceived relevance, received largely positive feedback. Several participants appreciated the depth and range of the modules, with specific praise for "Be Kind and Relax," which was the most frequently mentioned (n=6, 46%) as helpful.

Other modules, including "Be Thankful," "Look After Your Body," and those incorporating breathing and journaling exercises, were also described as useful, especially when they felt practical or easily incorporated into daily routines. However, some feedback suggested that the app's content was not always relevant for SPs (n=2, 15%):

> *A lot of the things are targeted at a person with diabetes, and I'm the support person.* [SP (father), 51 years, NZ]

Several participants (n=3, 23%) expressed personal preferences or limitations that influenced how helpful they found certain modules—for example, difficulty engaging with meditation due to chronic pain or discomfort with emotional content:

> *Breathing or meditation, like that's not for me.* [Adolescent and young adult, 22 years, United States]
>
> *Certain modules are kind of fit for different personality types.* [Adolescent and young adult, 20 years, United States]

Two felt that the tone or framing of certain concepts could be off-putting:

XSL•FO
RenderX

> *Be Thankful...felt a little suggestive of "hey, it could be worse."* [Adolescent and young adult, 21 years, United States]

User Perspectives: Recommendations for Future Development

Participants shared a range of suggestions for how the app could be enhanced in future iterations. A central area of focus was the well-being tree, which most participants (n=9, 69%) valued but wished could evolve further to include customizable or seasonal elements, new objects (eg, flowers or gardens), and even motivational consequences such as wilting:

> *I can't let my digital tree die, yeah. So, I think that would be a good motivator.* [Adolescent and young adult, 22 years, NZ]

Some participants (n=4, 31%) also proposed alternatives to the tree, including growing spaceships or gardens, to support greater personal relevance and customizability.

Some participants (n=3, 23%) also highlighted opportunities to integrate the app with existing digital tools and networks:

> *It would be really cool to see this kind of app and the diabetes tracking combined with one of the health watches...to see how my blood sugar reacts to specific things.* [Adolescent and young adult, 24 years, United States]

> *if you want to make it New Zealand specific you could always integrate it with a lot of the community groups we have like Diabetes Youth Auckland.* [Adolescent and young adult, 20 years, NZ]

Personalization was another interest. Overall, 2 (15%) users requested the ability to tailor the app to their age, stage of life, and content preferences:

> *I would make it have settings for specific age groups/stages of life for a more personable experience.* [Adolescent and young adult, 21 years, United States]

> *I would give different options for voices and tones.* [Adolescent and young adult, 20 years, United States]

SPs (n=3, 23%) called for more targeted content and pathways:

> *Maybe there could be like a category around like parents or friends or something.* [SP (friend), 16 years, NZ]

Some participants (n=3, 23%) also saw potential in incorporating more lived experience content to foster connection and normalize challenges:

> *The lived experiences part is cool...maybe a moderated submission element where people could even anonymously post lived experiences would be great.* [Adolescent and young adult, 20 years, NZ]

Others envisioned video or podcast libraries featuring community members and health professionals. Finally, some participants (n=5, 38%) wanted more interactivity to maintain engagement:

> *More activities in between the words could be more engaging.* [Adolescent and young adult, 21 years, United States]

> *you could almost have...friend quests...like seeing each other's trees.* [SP (father), 51 years, NZ]

User Perspectives: Impact of the Lift App

Many participants (n=7, 54%) described the app as having a meaningful emotional or behavioral impact, particularly in promoting connection, self-awareness, and practical coping strategies. For 1 (8%) adolescent and young adult and SP dyad, the shared experience of using the app strengthened their sense of support:

> *It was a really fun experience to be able to actually do something together...kind of like Duo Lingo...stuff you can do daily.* [SP (father), 54 years, NZ]

Others appreciated the app's emphasis on emotional reflection and noted the capacity for building longer-term skills:

> *It made me more self-aware of my emotions and inspired me to control what I can to change those feelings.* [SP (partner), 22 years, NZ]

> *I liked how it allowed me to build better strategies to cope with having diabetes.* [Adolescent and young adult, 20 years, NZ]

One participant also highlighted a sense of common humanity derived from the app:

> *It was nice to hear from other people with diabetes because it makes you feel less alone.* [Adolescent and young adult, 16 years, NZ]

Discussion

Principal Findings

This qualitative study builds on a recently published feasibility evaluation of the *Lift* digital health intervention for adolescents and young adults with T1D, which demonstrated promising preliminary findings and showed that *Lift* is safe, acceptable, and engaging among users [22]. This study aimed to capture user experiences, perceived impact, and suggestions for improving future iterations of the app. Overall, participants in this study described *Lift* as engaging, easy to use, and impactful and made suggestions for future improvements, including personalization and greater gamification.

Participants consistently expressed a desire for content tailored to their age, role (eg, SP vs young person), and personal preferences (eg, voice, pace, tone, and interactivity). This aligns with existing research showing that personalization can enhance engagement with digital health interventions by increasing the perceived relevance of content, thereby improving user motivation [29]. However, evidence also cautions that misaligned content, when not well matched to user needs or preferences, can lead to disengagement [30]. Indeed, this was echoed by some of the parent participants, who felt that they were not the "target market" for certain content, particularly content specific to diabetes.

XSL•FO
RenderX

For self-guided digital health interventions such as *Lift*, fostering engagement requires balancing broad appeal with delivering evidence-based, specific, and useful content. Moreover, it is difficult to predict the specific preferences and characteristics of end users, suggesting that future iterations of *Lift* and similar digital interventions could benefit from modular and adaptive content pathways that allow users to autonomously tailor their experience (such as content targeted to parents of adolescents and young adults living with diabetes). Indeed, this aligns with the self-determination theory, which, in digital health contexts, postulates that autonomy and perceived relevance are essential for sustaining intrinsic motivation to engage [31].

Qualitative data collected in this study highlighted that participants appreciated shared experiences (eg, stories, lived experience videos, and quotes), and many proposed interactive or social features (eg, forums, friend quests, and community groups) as ways to further improve the app. Participants highlighted that this content reduced feelings of isolation and enhanced the impact of *Lift*. These insights can be contextualized within some of the core frameworks underpinning the app, including self-compassion and acceptance and commitment therapy, which emphasize the role of common humanity and connectedness as buffers against shame and stress, particularly in diabetes [32]. This is a particularly impactful finding when taking into consideration the high levels of self-stigma and shame in diabetes [33].

Research supports social connection as a predictor of positive coping and a powerful mediator of diabetes distress [34,35], and previous research highlights that youth with diabetes want peer support features in digital interventions [36]. However, what remains unclear in the context of digital health is how to sustainably establish safe and moderated online spaces, particularly for users who are minors. At least initially, further work could be done to establish paired interactive features (such as “friend quests” and challenges), and future development could explore how platforms for community contribution or interaction could safely and sustainably be established, thus further supporting user retention, engagement, and the provision of connection and common humanity.

In keeping with findings regarding the role of social features in incentivizing engagement, interactive features were also an area of rich data in this study. Participants noted that app features such as the well-being tree, check-ins, and progress indicators were viewed as motivating; however, these needed more variety and novelty to sustain engagement. Many participants viewed these features as not just fun but also as motivational tools that helped them re-engage with the app and their self-care routines.

These findings can be contextualized within behavioral design theory, particularly through the value of design components such as feedback loops and progress indicators, which can increase engagement by satisfying the psychological need for competence and achievement [37]. For instance, the idea of the tree wilting when not used (and flourishing when attended to) mirrors habit loop mechanics, in which behavior is reinforced by visible progress and mild consequences for inaction. This is consistent with other digital health research, which has found that gamification increases engagement and facilitates greater behavior change [38,39]. It stands to reason that increasing gamification within *Lift* may increase and sustain both motivation and engagement through the provision of small, meaningful rewards.

There are also limitations to this study. It was a small qualitative study with a relatively homogeneous group of participants; therefore, generalized conclusions cannot be made. The participants who volunteered for this study after completing the *Lift* feasibility study may also represent more engaged users. Finally, this study did not explore cultural and health care system differences between the 2 study sites (NZ and United States), which could have affected the overall findings.

Conclusions

This qualitative study highlights user experiences and key benefits of *Lift* from the perspective of adolescents and young adults living with diabetes and their SPs. Participants reported improved emotional awareness, stress reduction, and reflective practices attributed to the app. These findings reinforce the value of emotionally intelligent, user-driven design in digital health. As adolescents and young adults and their support networks continue to navigate the complexities of living with diabetes, digital tools such as *Lift* show promise not just in promoting self-care, but also in fostering connection, agency, and a sense of shared humanity. This study also offers key insights both for future iterations of the *Lift* app and for digital health interventions more broadly. Future research should explore how such tools can be scaled, personalized, and embedded into everyday care, turning small digital moments into meaningful, sustained impact.

Acknowledgments

The authors thank the adolescents and the young adults and their support people who participated in this study.

Funding

This project was funded by a Breakthrough T1D project grant (grant 3-SRA-2023-1279-M-B).

Data Availability

Due to the small sample size, which may make participants potentially identifiable, as well as considerations regarding Indigenous data sovereignty, the data are not available.

XSL•FO
RenderX

Authors' Contributions

AS came up with the concept and design of the study and is the principal investigator on the Breakthrough T1D project grant. KMB recruited participants for the qualitative study with assistance of CH. JLS led the US arm of the study. CJ, JM, RP, HT, and MdB provided clinical oversight and supported recruitment in New Zealand. ALB conducted the focus groups and interviews with the assistance of KMB. KHG and KMB conducted the qualitative analysis with support from ALB, CH, and AS. AC provided statistical support. KMB wrote the initial draft of the manuscript. All authors edited and approved the final versions of the manuscript.

Conflicts of Interest

The *Lift* app was developed by AS and the research team at the University of Auckland and is a not-for-profit initiative. Author AS is an Associate Editor of *JMIR Pediatrics and Parenting.* All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Table for the interview schedule for adolescents and young adults and their caregivers.
[PDF File, 88 KB - diabetes_v11i1e79896_app1.pdf]

Checklist 1
COREQ Checklist
[PDF File, 90 KB - diabetes_v11i1e79896_app2.pdf]

References

1. Morrissey EC, Casey B, Dinneen SF, Lowry M, Byrne M. Diabetes distress in adolescents and young adults living with type 1 diabetes. Can J Diabetes 2020 Aug;44(6):537-540. [doi: 10.1016/j.jcjd.2020.03.001] [Medline: 32507646]
2. Stahl-Pehe A, Glaubitz L, Bächle C, et al. Diabetes distress in young adults with early-onset type 1 diabetes and its prospective relationship with HbA1c and health status. Diabet Med 2019 Jul;36(7):836-846. [doi: 10.1111/dme.13931] [Medline: 30761589]
3. Collins MM, Corcoran P, Perry IJ. Anxiety and depression symptoms in patients with diabetes. Diabet Med 2009 Feb;26(2):153-161. [doi: 10.1111/j.1464-5491.2008.02648.x] [Medline: 19236618]
4. Hanson CL, De Guire MJ, Schinkel AM, Kolterman OG, Goodman JP, Buckingham BA. Self-care behaviors in insulin-dependent diabetes: evaluative tools and their associations with glycemic control. J Pediatr Psychol 1996 Aug;21(4):467-482. [doi: 10.1093/jpepsy/21.4.467] [Medline: 8863457]
5. Kakleas K, Kandyla B, Karayianni C, Karavanaki K. Psychosocial problems in adolescents with type 1 diabetes mellitus. Diabetes Metab 2009 Nov;35(5):339-350. [doi: 10.1016/j.diabet.2009.05.002] [Medline: 19700362]
6. Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989. [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
7. McCoy RG, Kidney RS, Holznagel D, Peters T, Madzura V. Challenges for younger adults with diabetes. Minn Med 2019;102(2):34-36. [Medline: 31889734]
8. Pyatak EA, Sequeira PA, Whittemore R, Vigen CP, Peters AL, Weigensberg MJ. Challenges contributing to disrupted transition from paediatric to adult diabetes care in young adults with type 1 diabetes. Diabet Med 2014 Dec;31(12):1615-1624. [doi: 10.1111/dme.12485] [Medline: 24798586]
9. Shulman R, Luo J, Shah BR. Mental health visits and low socio-economic status in adolescence are associated with complications of type 1 diabetes in early adulthood: a population-based cohort study. Diabet Med 2018 Jul;35(7):920-928. [doi: 10.1111/dme.13633] [Medline: 29608218]
10. Nakhla M, Shulman R, Dimeglio L. Mental health matters: limited support remains a barrier to optimal care for youth with diabetes. Can J Diabetes 2021 Jul;45(5):379-380. [doi: 10.1016/j.jcjd.2021.05.003] [Medline: 34176609]
11. Cao VT, Anderson BJ, Eshtehardi SS, et al. “We are a family with diabetes”: parent perspectives on siblings of youth with type 1 diabetes. Fam Syst Health 2021 Jun;39(2):306-315. [doi: 10.1037/fsh0000612] [Medline: 34410773]
12. Chan KK, Shorey S. Experiences and needs of children with siblings diagnosed with type 1 diabetes: a mixed studies systematic review. J Pediatr Nurs 2022;63:1-8. [doi: 10.1016/j.pedn.2021.12.002] [Medline: 34929508]
13. Whittemore R, Jaser S, Chao A, Jang M, Grey M. Psychological experience of parents of children with type 1 diabetes: a systematic mixed-studies review. Diabetes Educ 2012;38(4):562-579. [doi: 10.1177/0145721712445216] [Medline: 22581804]
14. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
15. Bernstein CM, Stockwell MS, Gallagher MP, Rosenthal SL, Soren K. Mental health issues in adolescents and young adults with type 1 diabetes: prevalence and impact on glycemic control. Clin Pediatr (Phila) 2013 Jan;52(1):10-15. [doi: 10.1177/0009922812459950] [Medline: 22988007]

XSL•FO
RenderX

16. Garner K, Boggiss A, Jefferies C, Serlachius A. Digital health interventions for improving mental health outcomes and wellbeing for youth with type 1 diabetes: a systematic review. Pediatr Diabetes 2022 Mar;23(2):258-269. [doi: 10.1111/pedi.13304] [Medline: 34913548]
17. Maharaj A, Lim D, Murphy R, Serlachius A. Comparing two commercially available diabetes apps to explore challenges in user engagement: randomized controlled feasibility study. JMIR Form Res 2021 Jun 16;5(6):e25151. [doi: 10.2196/25151] [Medline: 34132640]
18. Butler S, Sculley D, Santos DS, et al. Usability of eHealth and mobile health interventions by young people living with juvenile idiopathic arthritis: systematic review. JMIR Pediatr Parent 2020 Dec 1;3(2):e15833. [doi: 10.2196/15833] [Medline: 33258786]
19. Bevan Jones R, Stallard P, Agha SS, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940. [doi: 10.1111/jcpp.13258] [Medline: 32572961]
20. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: a scoping review. Digit Health 2023;9:20552076231219117. [doi: 10.1177/20552076231219117] [Medline: 38107986]
21. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
22. Serlachius A, McClintock J, Boggiss A, et al. Lifting the wellbeing of adolescents and young adults with type 1 diabetes: a feasibility study of the LIFT app. Int J Med Inform 2025 Nov;203:105992. [doi: 10.1016/j.ijmedinf.2025.105992] [Medline: 40446520]
23. Thabrew H, Boggiss AL, Lim D, et al. Well-being app to support young people during the COVID-19 pandemic: randomised controlled trial. BMJ Open 2022 May 19;12:e058144. [doi: 10.1136/bmjopen-2021-058144]
24. Garner K, Thabrew H, Lim D, Hofman P, Jefferies C, Serlachius A. Exploring the usability and acceptability of a well-being app for adolescents living with type 1 diabetes: qualitative study. JMIR Pediatr Parent 2023 Dec 22;6:e52364. [doi: 10.2196/52364] [Medline: 38153804]
25. Serlachius A, Schache K, Boggiss A, et al. Coping skills mobile app to support the emotional well-being of young people during the COVID-19 pandemic: protocol for a mixed methods study. JMIR Res Protoc 2020 Oct 8;9(10):e23716. [doi: 10.2196/23716] [Medline: 32991303]
26. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
27. Couch D. Realist analysis. In: Okoko JM, Tunison S, Walker KD, editors. Varieties of Qualitative Research Methods: Selected Contextual Perspectives: Springer; 2023:407-412.
28. Harrison Ginsberg K, Babbott K, Serlachius A. Exploring participants' experiences of digital health interventions with qualitative methods: guidance for researchers. J Med Internet Res 2024 Nov 28;26:e62761. [doi: 10.2196/62761] [Medline: 39607999]
29. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015;1:2055207615595335. [doi: 10.1177/2055207615595335] [Medline: 29942544]
30. Updegraff JA, Sherman DK, Luyster FS, Mann TL. The effects of message quality and congruency on perceptions of tailored health communications. J Exp Soc Psychol 2007;43(2):249-257. [doi: 10.1016/j.jesp.2006.01.007] [Medline: 18958299]
31. Şener B, Umulu S, Yilmaz AO. Digital health interventions for promoting sleep wellbeing: a design approach using self-determination theory. In: Bohemia E, Buck L, Grierson H, editors. Proceedings of the 24th International Conference on Engineering and Product Design Education (E&PDE 2022): The Design Society; 2022. [doi: 10.35199/EPDE.2022.55]
32. Onu DU, Obi-keguna CN, Oguguam ON, Ajaero CK, Igwe EJ. Social support may buffer, to an extent, the impact of stigma on health-related quality of life among type 2 diabetes mellitus patients. Discov Public Health 2025;22(1):29. [doi: 10.1186/s12982-025-00415-x]
33. Speight J, Holmes-Truscott E, Garza M, et al. Bringing an end to diabetes stigma and discrimination: an international consensus statement on evidence and recommendations. Lancet Diabetes Endocrinol 2024 Jan;12(1):61-82. [doi: 10.1016/S2213-8587(23)00347-9] [Medline: 38128969]
34. Chan CK, Cockshaw W, Smith K, Holmes-Truscott E, Pouwer F, Speight J. Social support and self-care outcomes in adults with diabetes: the mediating effects of self-efficacy and diabetes distress. Results of the second diabetes MILES - Australia (MILES-2) study. Diabetes Res Clin Pract 2020 Aug;166:108314. [doi: 10.1016/j.diabres.2020.108314] [Medline: 32653506]
35. Luo D, Cai X, Wang H, Wang Y, Xu J. The role of peer social relationships in psychological distress and quality of life among adolescents with type 1 diabetes mellitus: a longitudinal study. BMC Psychiatry 2024 Apr 11;24(1):270. [doi: 10.1186/s12888-024-05692-5] [Medline: 38605327]
36. Boggiss AL, Consedine NS, Schache KR, et al. Exploring the views of adolescents with type 1 diabetes on digital mental health interventions: what functionality and content do they want? Diabet Med 2021 Nov;38(11):e14591. [doi: 10.1111/dme.14591] [Medline: 33930211]
37. Fogg B. A behavior model for persuasive design. In: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology: Association for Computing Machinery; 2009.

XSL•FO
RenderX

38. Berglund A, Jaarsma T, Berglund E, Strömberg A, Klompstra L. Understanding and assessing gamification in digital healthcare interventions for patients with cardiovascular disease. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):630-638. [doi: 10.1093/eurjcn/zvac048] [Medline: 35709297]
39. King D, Greaves F, Exeter C, Darzi A. “Gamification”: influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
Lift: Lift: Thriving with Diabetes
NZ: New Zealand
SP: support person
T1D: type 1 diabetes
uMARS: end-user version of the Mobile Application Rating Scale

Edited by I Steenstra; submitted 03.Jul.2025; peer-reviewed by H Witteman, N Peterson; revised version received 26.Feb.2026; accepted 30.Mar.2026; published 28.May.2026.

Please cite as:
Babbott KM, Ginsberg KH, Boggiss AL, McClintock J, Hudson C, Hofman P, Paul R, Cavadino A, Thabrew H, Jefferies C, de Bock M, Sherr JL, Serlachius A
Exploring User Experiences With the Lift App for Emotional Well-Being Among Youth With Type 1 Diabetes: Qualitative Study
JMIR Diabetes 2026;11:e79896
URL: https://diabetes.jmir.org/2026/1/e79896
doi:10.2196/79896

© Katie M Babbott, Kristin Harrison Ginsberg, Anna L Boggiss, Joanna McClintock, Chloe Hudson, Paul Hofman, Ryan Paul, Alana Cavadino, Hiran Thabrew, Craig Jefferies, Martin de Bock, Jennifer L Sherr, Anna Serlachius. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 28.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Serious Digital Game (SugarVita) to Support Diabetes Self-Management: Pilot Randomized Controlled Trial

Edouard Reinders[1,2,3], MD; Elma Cruts - Van de Meulengraaf[4]; Ryan De Vries[3], MSc; Natal van Riel[3], Prof Dr; Pieter Van Gorp[3], Prof Dr; Jeanne Dieleman[5], PhD; Pleun Wouters - van Poppel[4], PhD; Harm Haak[2], Prof Dr

[1]Internal Medicine, Maastricht University Medical Centre, Maastricht, Limburg, The Netherlands
[2]Care and Public Health Research Institute, Maastricht, Limburg, The Netherlands
[3]Department of Biomedical Engineering, Eindhoven University of Technology, Het Eeuwsel 53, Eindhoven, North Brabant, The Netherlands
[4]Internal Medicine, Máxima Medisch Centrum, Veldhoven, North Brabant, The Netherlands
[5]Science Office, Máxima Medical Center, Veldhoven, The Netherlands

Corresponding Author:
Edouard Reinders, MD
Internal Medicine, Maastricht University Medical Centre, Maastricht, Limburg, The Netherlands

Abstract

Background: Serious digital games have been proposed as a novel approach to support diabetes education and self-management, but evidence regarding their effectiveness remains limited.

Objective: This study aimed to evaluate the effects of SugarVita, a serious game for people with type 2 diabetes, on diabetes-related knowledge, self-confidence, and self-management. Secondary outcomes included hemoglobin A_{1c} (HbA_{1c}), engagement, and user evaluation.

Methods: In this pilot randomized controlled trial, 30 adults with type 2 diabetes were randomized to SugarVita plus standard care or standard care alone for 8 weeks. Outcomes were assessed before and after the intervention using validated questionnaires and laboratory HbA_{1c} values. Within-group changes were analyzed using Wilcoxon signed-rank tests and between-group differences using Mann-Whitney *U* tests. Bonferroni correction was applied for multiple primary outcomes.

Results: No statistically significant between-group differences were observed for diabetes-related knowledge, self-confidence, or self-management after correction for multiple testing. Both groups showed numerical improvements over time. HbA_{1c} decreased significantly within the intervention group (median 73.0, IQR 70.8-81.5 to median 64.5, IQR 60.8-72.0 mmol/mol; P=.007), whereas no significant change was observed in the control group. Greater total playtime was moderately associated with HbA_{1c} reduction. User evaluations indicated high perceived educational value.

Conclusions: Participants reported positive experiences with SugarVita and perceived the game as educational and user-friendly. No statistically significant between-group differences were observed for the primary outcomes. These findings support the feasibility and acceptability of SugarVita as a digital educational intervention and warrant further evaluation in larger studies.

Trial Registration: ClinicalTrials.gov NCT06392178; https://clinicaltrials.gov/study/NCT06392178

(JMIR Diabetes 2026;11:e99345) doi:10.2196/99345

KEYWORDS

diabetes self-management and education support; digital education; serious games; eHealth; mobile app

Introduction

The global incidence of diabetes is on the rise, with projections estimating that nearly 580 million people worldwide will be affected by 2030 [1]. Traditionally, type 2 diabetes mellitus (T2DM) has been considered a disease that develops later in life. However, an increasing number of younger individuals are now being diagnosed [2]. Diabetes is associated with early and late complications that, despite advances in cardiovascular risk management, continue to drive an elevated risk of cardiovascular events [3]. Consequently, individuals with T2DM have higher mortality rates and a shorter life expectancy than those without the disease [4].

Effective disease management necessitates that patients understand the factors influencing blood glucose levels, such as diet, physical activity, and mental well-being [5]. Lifestyle changes, including healthier choices in daily routines, are often essential to achieve good regulation [6]. Better disease control results in fewer day-to-day complaints, a reduced risk of long-term complications, and improved life expectancy [7]. Therefore, early awareness of the disease, its contributing factors, and its consequences is crucial. Guidance from health

XSL•FO
RenderX

care professionals, including general practitioners, internists, and specialized nurses, is vital in supporting this process. Traditionally, patient education has involved one-on-one sessions and ongoing counseling provided by these professionals.

Given the increasing prevalence of diabetes, the growing demand for health care professionals, and the associated medical costs, innovative solutions are required to ensure sustainable care. Current diabetes guidelines emphasize the use of digital tools for patient education [8]. The advantage of these tools, such as smartphones, lies in their widespread availability and easy access to diabetes-related apps through platforms such as the Google Play Store and the iOS Store. With the global proliferation of smartphones, these tools offer a promising means of providing diabetes education.

One innovative approach involves using educational games to teach patients about their disease in an engaging manner. These games aim to combine entertainment with education, with the intention of enhancing engagement and supporting learning compared to traditional one-on-one sessions. In recent years, there has been a surge in the development of digital diabetes games, as evidenced by their growing presence in both scientific literature and app stores [9]. SugarVita, a digital educational game, was developed from this perspective [10,11]. It aims to assist individuals with T2DM in managing their condition more effectively, strengthening their confidence in daily diabetes care, and improving their understanding of the disease.

Although interest in digital diabetes games has increased in recent years, evidence regarding their effectiveness remains limited. Previous studies have often focused on feasibility, usability, or user satisfaction, while effects on diabetes-related knowledge, self-management, and clinical outcomes have been evaluated less frequently. Therefore, additional research is needed to better understand the role of serious games in supporting diabetes education and self-management in daily practice [9].

The primary objective of this study was to evaluate the impact of playing SugarVita on diabetes-related knowledge, self-management, and self-confidence in individuals with T2DM, as these outcomes were considered the most direct measures of the educational objectives of the intervention. Hemoglobin A_{1c} (HbA_{1c}) was included as a secondary outcome to explore whether improvements in diabetes education and self-management might translate into changes in glycemic control. Additional secondary objectives included analyzing player engagement data (eg, total playtime) and participant evaluations of the game (eg, perceived enjoyment and educational value).

Methods

Study Design and Participants

An 8-week randomized controlled trial was conducted between June 2024 and December 2024 at the Department of Internal Medicine, Máxima Medical Center, Eindhoven and Veldhoven, the Netherlands. The study was designed as a pilot randomized controlled trial including 30 participants to explore its impact on clinical outcomes and user experience.

All participants were recruited from the outpatient clinic of the Máxima Medical Center. Inclusion criteria were as follows: (1) a diagnosis of T2DM, (2) HbA_{1c} levels >64 mmol/mol at baseline, (3) BMI >25 kg/m^2, and (4) aged ≥18 years. Exclusion criteria included: (1) pregnancy or breastfeeding, (2) ongoing treatment for malignancy, and (3) inability to speak Dutch.

During the first visit, participants signed informed consent forms and were randomized to either the intervention or control group. Randomization was stratified by age and baseline HbA_{1c} values. Participants were randomized using block randomization with sealed allocation lots prepared before the start of participant enrollment. Eligible participants were assigned to the intervention or control group only after completion of baseline procedures. Allocation was concealed until the moment of assignment. A brief overview of their medical history and current medication use was registered. In addition, changes in glucose-lowering medication occurring shortly before or during the study period were recorded and reviewed during follow-up. Additionally, HbA_{1c} levels from the participants' most recent routine clinical visits were documented. All participants completed 3 questionnaires at baseline and again at the end of the 8-week study period, assessing diabetes self-confidence, self-management, and diabetes-related knowledge. The same questionnaires were administered at baseline and follow-up. As participants did not receive feedback on their responses, any practice effects due to repeated testing were expected to be limited and equally applicable to both study groups.

Participants in the control group did not play SugarVita and received standard care according to their individual treatment plans during the 8-week follow-up. Participants in the intervention group were instructed to download and play SugarVita on their own smartphone, tablet, or PC from the Google Play Store (Google LLC) or the Apple App Store (Apple Inc). Tablets (Samsung Galaxy Tab A6, Samsung Electronics) were provided by the research team if needed. After 1 week, participants of the intervention group received a follow-up phone call to address any technical issues related to logging into or using SugarVita. No additional educational or behavioral support was provided. At the start of the study, participants in the intervention group also received a paper-based assignment card to be completed over the 8-week study period. These challenges were designed to enhance engagement and motivate participants to continue playing the game throughout the study period. For example, in the fourth week, participants were asked to play the game with a relative or friend in a multiplayer setting (refer to Multimedia Appendix 1 for the assignment card).

During the second visit, at the end of the study, all participants completed the same 3 questionnaires again. Additionally, participants in the intervention group filled out an extra questionnaire evaluating their experience with SugarVita (Multimedia Appendix 2).

Standard Diabetes Care

All participants received standard diabetes care in accordance with Dutch national guidelines for T2DM management, as

XSL•FO
RenderX

outlined in the NHG-Standaard Diabetes mellitus type 2 [12]. Standard care consisted of regular outpatient follow-up with an internist and/or diabetes nurse specialist. During these consultations, lifestyle-related topics such as dietary habits, physical activity, weight management, medication adherence, and self-monitoring of blood glucose were routinely discussed.

Diabetes education was primarily delivered through individual consultations rather than structured group-based education programs. Participants also had access to written educational materials provided by the hospital or national diabetes organizations, covering general aspects of diabetes management and lifestyle modification.

No structured digital educational tools, mobile health apps, or serious games were part of routine care during the study period. Both the intervention and control groups received the same standard care. No additional educational sessions were introduced as part of the study protocol, apart from access to the SugarVita serious game for participants randomized to the intervention group.

Ethical Considerations

The study was registered at ClinicalTrials.gov (NCT06392178) and conducted in accordance with the Declaration of Helsinki. The Medical Research Ethics Committee of Máxima Medical Center reviewed the study and determined that it does not fall under the scope of the Dutch Medical Research Involving Human Subjects Act (N24.020); therefore, formal ethics approval was not required. Eligible participants provided written informed consent prior to participation.

Outcome Measures

Self-Confidence

Self-confidence in diabetes self-care was assessed using the validated Dutch questionnaire “Vertrouwen in Diabetes Zelfzorg” [13]. This instrument consists of 21 questions in which participants rate their confidence in performing various self-care tasks. Scores range from 21 to 105, with higher scores indicating greater self-confidence (Multimedia Appendix 3).

Self-Management

Self-management was assessed using the validated Dutch version of the Diabetes Management Self-Efficacy Scale [14]. This questionnaire consists of 20 items in which participants estimate their perceived ability to perform various tasks related to diabetes management. Total scores range from 20 to 100, with higher scores indicating greater perceived self-management capacity (Multimedia Appendix 3).

Diabetes-Related Knowledge

Diabetes-related knowledge was assessed using a translated version of the Diabetes Knowledge Test [15]. As no validated Dutch instrument was available at the time of the study, the original English version was translated into Dutch. The resulting questionnaire consists of 19 items related to diabetes knowledge. Total scores range from 0 to 59, with a maximum possible score of 59. The Dutch version used in this study has not yet been validated (Multimedia Appendix 3).

Hemoglobin A_{1c}

HbA_{1c} values were obtained from routine clinical measurements documented in the electronic health record. No additional blood sampling was required. Given the duration of the intervention (8 weeks) and the expected delay before HbA_{1c} reflects sustained glycemic changes, postintervention values were included if measured within 12 weeks after study completion. Baseline HbA_{1c} values were accepted if collected within 12 weeks prior to study enrollment.

Player Data

Player activity was monitored remotely throughout the intervention period. Each participant was assigned an anonymous user account to access SugarVita. This enabled the research team to track detailed gameplay data, including login frequency, number of game sessions, session completion rates, total playtime, and in-game behavior. Gameplay data also included the number and accuracy of diabetes-related multiple-choice questions triggered during gameplay, the gameplay mode (solo vs multiplayer), and device type. Gameplay data were linked to study identifiers rather than personal information, stored securely, and accessible only to members of the study team.

Evaluation of SugarVita

Upon completion of the intervention, participants in the intervention group were invited to fill out an additional evaluation questionnaire regarding their experience with SugarVita. The questionnaire consisted of 8 items rated on a 10-point Likert scale (1=strongly disagree and 10=strongly agree), assessing aspects such as perceived educational value and level of challenge. Participants were also given the opportunity to provide open-ended comments to elaborate their responses.

SugarVita

SugarVita was developed at our outpatient clinic in collaboration with Eindhoven University of Technology and in consultation with a group of patients with diabetes who volunteered for this project and did not participate in the current study. Various game types, including quiz and puzzle games, were tested, but patients preferred a digital version of a classic board game. The development process and selection criteria are described in a previously published article [11]. The core of SugarVita is based on the Eindhoven Diabetes Education Simulator, a physiology-based mathematical model that predicts blood glucose responses in a semipersonalized manner over 24 hours. The simulator incorporates key factors influencing glycemic control, allowing the game to simulate realistic glucose dynamics in response to player decisions [16].

SugarVita resembles the classic *Game of the Goose* or *Snakes and Ladders* game, with a board consisting of tiles and players rolling a digital die to advance. SugarVita can be played alone, with friends or family in multiplayer mode, or against a computer opponent. The game starts in the morning with medication intake and breakfast, then progresses through the day, with each tile representing 20 minutes of the day. When landing on a tile, players make choices regarding everyday

XSL•FO
RenderX

events (eg, exercising, eating, snacking, and commuting). The player's blood glucose level is continuously visible during the game and fluctuates based on their actions. Players earn the most points by maintaining their blood glucose levels within a certain range. Additionally, some tiles present random multiple-choice knowledge questions about diabetes, with correct answers earning extra points. The player with the most points at the end of the day wins the game. At the end of the virtual day, players receive a summary of their performance, including their glucose curve. Animated pawns (eg, a piece of fruit) are used to navigate the board, and completing certain challenges unlocks new pawns (refer to Figure 1 for an illustrative screenshot).

Figure 1. Screenshot of SugarVita.

Statistical Analysis

Within-group changes from baseline to follow-up were assessed using Wilcoxon signed-rank tests. Between-group differences in outcome scores (before and after) were evaluated using the Mann-Whitney *U* test. Given the evaluation of 3 primary outcomes, a Bonferroni correction was applied to diabetes-related knowledge, self-confidence, and self-management, resulting in a significance threshold of $P<.017$. Results are reported as medians with IQRs and associated *P* values. A sensitivity analysis was performed using a narrower follow-up window by excluding participants with follow-up HbA_{1c} measurements obtained within 6 weeks after study inclusion. Associations between total playtime and changes in clinical outcomes were explored using Pearson correlation coefficients. A 2-sided *P* value of <.05 was considered to indicate a possible association in these exploratory analyses.

All statistical analyses were performed using SPSS software (version 27; IBM Corp).

Results

Demographics

A total of 30 participants were included, of whom 16 (53.3%) were women. The mean age was 61.4 (SD 9.0) years, and the mean BMI was 32.7 (SD 6.9) kg/m². Mean baseline HbA_{1c} was 76.3 (SD 12.6) mmol/mol, and the mean duration of diabetes was 14.8 (SD 8.9; range 3 - 46) years. A total of 22 (73.3%) participants used insulin, with a mean dose of 0.60 (SD 0.47) U/kg. Most participants were treated with metformin (n=23, 76.7%), whereas fewer participants used glucagon-like peptide-1 (GLP-1) receptor agonists (n=18, 60%), sulfonylureas (n=14, 46.7%), or sodium-glucose co-transporter-2 (SGLT2) inhibitors (n=7, 23.3%). All 30 participants completed the study, and no relevant data were missing for the primary or secondary outcomes. Participant flow throughout the study is shown in (Figure 2). Baseline characteristics are presented in Table 1.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) 2025 flow diagram. The number of individuals assessed for eligibility prior to enrollment was not systematically recorded.

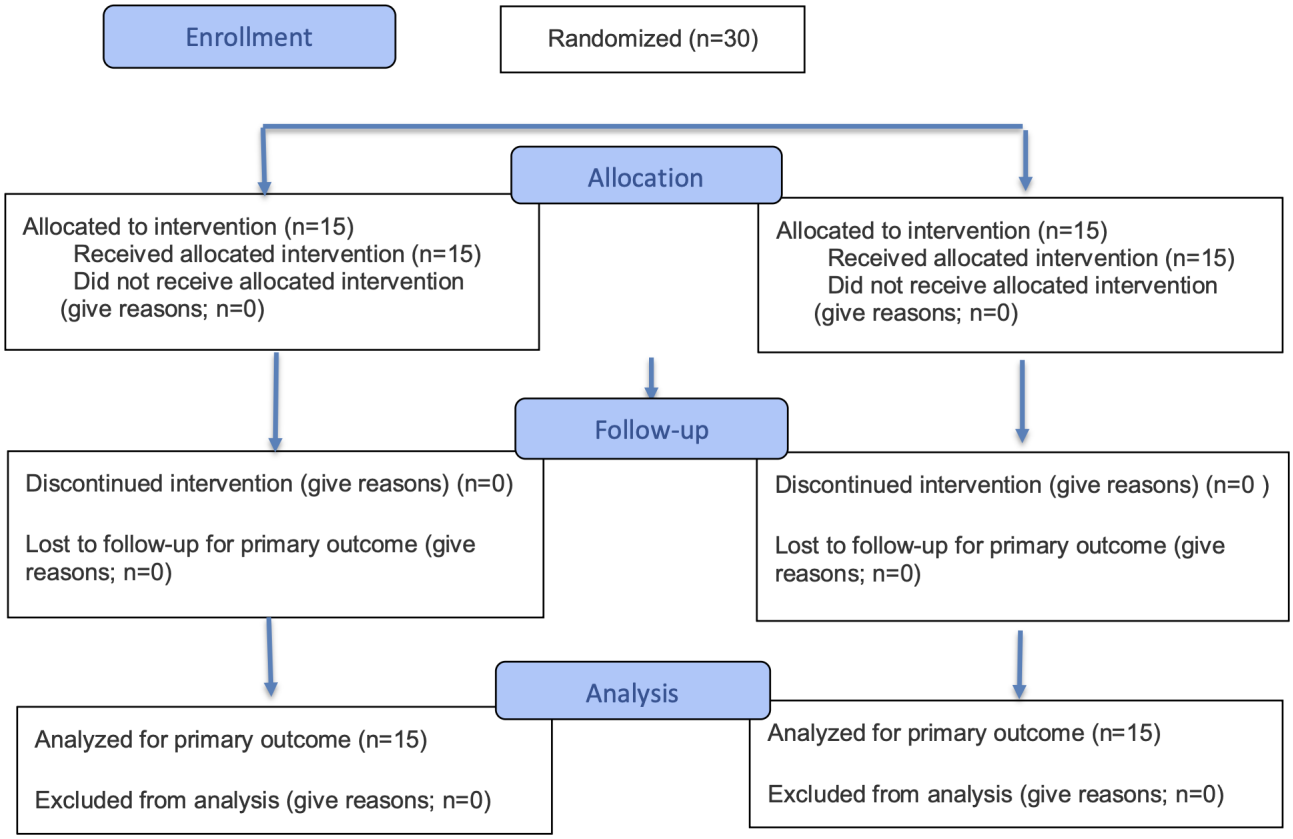


Table . Baseline participant characteristics (N=30).

	Total	Control group	Intervention group
Characteristic			
Female sex, n (%)	16 (53.3)	7 (46.7)	9 (60)
Age (years), mean (SD)	61.4 (9.0)	60.3 (10.4)	62.5 (7.5)
BMI (kg/m²), mean (SD)	32.7 (6.9)	29.6 (5.1)	35.9 (7.2)
HbA_{1c}[a] (mmol/mol), mean (SD)	76.3 (12.6)	75.4 (15.4)	77.2 (9.5)
HbA_{1c} (%), mean (SD)	9.1 (1.2)	9.0 (1.4)	9.2 (0.9)
Duration of diabetes (years), mean (SD)	14.8 (8.9)	12.4 (7.0)	17.3 (10.0)
Medication			
Insulin use, n (%)	22 (73.3)	11 (73.3)	11 (73.3)
Insulin dose (U/kg), mean (SD)[b]	0.60 (0.47)	0.67 (0.59)	0.52 (0.33)
Metformin use, n (%)	23 (76.7)	9 (60)	14 (93.3)
Sulfonylurea use, n (%)	14 (46.7)	9 (60)	5 (33.3)
GLP-1[c] receptor agonist use, n (%)	18 (60)	7 (46.7)	11 (73.3)
SGLT2[d] inhibitor use, n (%)	7 (23.3)	2 (13.3)	5 (33.3)

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]Insulin dose was calculated only for participants receiving insulin therapy.

[c]GLP-1: glucagon-like peptide-1.

[d]SGLT2: sodium-glucose co-transporter-2.

XSL•FO
RenderX

At baseline, participants in the intervention group had a higher BMI (mean 35.9, SD 7.2 vs mean 29.6, SD 5.1 kg/m²) and a longer duration of diabetes (mean 17.3, SD 10.0 vs mean 12.4, SD 7.0 years) compared to the control group. A greater proportion of participants in the intervention group were using metformin (n=14, 93.3% vs n=9, 60%), GLP-1 receptor agonists (n=11, 73.3% vs n=7, 46.7%), and SGLT2 inhibitors (n=5, 33.3% vs n=2, 13.3%). Medication changes occurring shortly before or during the study period were observed in 46.7% (n=7) of the participants in the intervention group and 53.3% (n=8) of the participants in the control group. Both treatment intensifications and deintensifications occurred in each group without a clear imbalance between study arms.

Primary Outcomes

Diabetes knowledge scores increased in the intervention group from 30.0 to 36.5 but not in the control group (31.0 to 31.0). The within-group change in the intervention group was not statistically significant (*P*=.07), nor was the between-group difference (*P*=.25). Self-confidence improved from 84.0 to 92.5 in the intervention group and from 90.0 to 96.0 in the control group, while self-management increased from 62.5 to 69.5 and from 69.0 to 76.0, respectively. None of the within-group or between-group comparisons for the primary outcomes reached statistical significance after Bonferroni correction (significance threshold *P*<.017; Table 2).

Table . Outcomes before and after the intervention.

Outcome	Control group				Intervention group				Between-group, *P* value
	Pre, median (IQR)	Post, median (IQR)	Change, median (IQR)	Within-group, *P* value	Pre, median (IQR)	Post, median (IQR)	Change, median (IQR)	Within-group, *P* value	
Diabetes knowledge	31.0 (26.0 to 33.5)	31.0 (24.5 to 36.5)	2.0 (−2.0 to 4.0)	.97	30.0 (25.8 to 36.2)	36.5 (33.0 to 38.5)	3.0 (0.5 to 10.5)	.07	.25
Self-confidence (VDZ[a])	90.0 (84.5 to 94.5)	96.0 (85.5 to 98.0)	3.0 (−0.5 to 5.5)	.08	84.0 (80.5 to 90.5)	92.5 (82.0 to 97.2)	6.0 (−0.5 to 8.5)	.07	.26
Self-management (DSMES[b])	69.0 (62.0 to 76.5)	76.0 (69.5 to 83.5)	5.0 (1.0 to 9.5)	.02	62.5 (56.8 to 80.5)	69.5 (63.0 to 79.0)	5.0 (1.0 to 11.0)	.07	.77
HbA_{1c}[c] (mmol/mol)	70.0 (66.5 to 83.5)	68.0 (63.5 to 73.5)	−5.0 (−12.5 to 2.5)	.08	73.0 (70.8 to 81.5)	64.5 (60.8 to 72.0)	−13.0 (−17.0 to −6.5)	.007	.25
HbA_{1c} (%)	8.6 (8.2 to 9.8)	8.4 (8.0 to 8.9)	−0.5 (−1.1 to 0.2)	.08	8.8 (8.7 to 9.6)	8.0 (7.7 to 8.8)	−1.2 (−1.6 to −0.6)	.007	.25

[a]VDZ: Vertrouwen in Diabetes Zelfzorg.

[b]DSMES: Diabetes Self-Management Education and Support.

[c]HbA_{1c}: hemoglobin A_{1c}.

Secondary Outcomes

No statistically significant between-group difference in HbA_{1c} change was observed (*P*=.25). Within-group analyses showed a significant HbA_{1c} reduction in the intervention group (73.0 to 64.5 mmol/mol; *P*=.007), whereas the smaller reduction in the control group was not statistically significant (70.0 to 68.0 mmol/mol; *P*=.08; Table 2). Baseline HbA_{1c} measurements were obtained a median of 31 (IQR 6-84) days before study inclusion, whereas follow-up HbA_{1c} measurements were obtained a median of 52 (IQR 10-82) days after inclusion. Sensitivity analysis using a narrower follow-up window yielded similar results, with median HbA_{1c} in the intervention group changing from 74.0 (IQR 70.8-85.5) to 64.5 (IQR 58.3-76.5) mmol/mol (*P*=.003).

Data are presented as median (IQR). Change scores represent postintervention minus preintervention values. *P* values represent within-group changes (Wilcoxon signed-rank test) and between-group comparisons of change scores (Mann-Whitney *U* test). The Bonferroni-corrected significance threshold for the primary outcomes was *P*<.0167.

Playtime

Among participants in the intervention group, the mean total playtime was 524 (SD 556) minutes over 8 weeks, corresponding to just over 1 hour of gameplay per week. Participants completed 387 game sessions during the intervention period. A total of 1081 diabetes-related knowledge questions were presented, of which 894 (82.7%) were answered correctly. A moderate negative correlation was found between longer playtime and greater HbA_{1c} reduction (Pearson *r*=−0.57; *P*=.07), suggesting a possible dose-response relationship; however, this association was not statistically significant. In contrast, only a weak and nonsignificant correlation was observed between playtime and improvement in self-management scores (Pearson *r*=0.22; *P*=.51; Table 3).

XSL•FO
RenderX

Table . SugarVita gameplay and participant evaluation[a].

Variable	Mean (SD; range)
Total minutes played	524.27 (555.57; 16.0-1734.0)
Total gameplay sessions	39.14 (48.39; 1.0-193.0)
Questions	
How enjoyable was it to play the game?	6.62 (2.18; 3.0-10.0)
How challenging was the game?	6.0 (2.35; 2.0-10.0)
How user-friendly was the game?	7.29 (1.9; 3.0-10.0)
Did you learn from the game?	8.14 (0.95; 7.0-10.0)
Would you recommend the game to others?	7.54 (2.11; 4.0-10.0)
Would you play the game more frequently?	5.25 (3.22; 1.0-10.0)
Do you think SugarVita contributes to diabetes-related knowledge?	8.04 (1.39; 5.0-10.0)
Do you think digital games are a useful support tool for people newly diagnosed with diabetes?	9.0 (1.0; 7.0-10.0)

[a]Descriptive statistics of gameplay activity and participant evaluation of the SugarVita serious game (intervention group only). Scores for evaluation items range from 1 (strongly disagree) to 10 (strongly agree).

Participant Evaluation of the Intervention

All 15 participants in the intervention group completed the evaluation questionnaire, and no item responses were missing. Overall evaluations of SugarVita were favorable across several domains, using a scale from 1 (strongly disagree) to 10 (strongly agree). Agreement with the statement that SugarVita may be a useful support tool for individuals newly diagnosed with diabetes received the highest score, with a mean of 9.0 (SD 1.0) out of 10. Perceived educational value also received high ratings, with a mean of 8.1 (SD 0.9) out of 10. Most of the participants expressed willingness to recommend the game to others (mean 7.5, SD 2.1) and rated it as user-friendly (mean 7.3, SD 1.9). Ratings for enjoyment and challenge were moderate, with mean scores of 6.6 (SD 2.2) and 6.0 (SD 2.3), respectively. Willingness to continue playing the game received a lower mean score of 5.25 (SD 3.22), suggesting room for improvement in long-term engagement (Table 3).

Discussion

Principal Findings

In this randomized controlled pilot study, no statistically significant between-group differences were observed for the primary outcomes after correction for multiple testing. Participants in the intervention group demonstrated a significant within-group reduction in HbA_{1c}, numerically higher diabetes knowledge scores, and positive evaluations of the game. However, given the absence of statistically significant between-group differences, these findings should be interpreted cautiously and cannot be attributed exclusively to the intervention. The observed changes likely reflect a combination of factors, including the intervention and routine diabetes care received during follow-up.

Serious games for diabetes self-management have been in development for several decades, with early examples targeting children with type 1 diabetes [17]. Over time, these games have evolved across various platforms, with mobile-based apps gaining popularity in recent years due to the widespread adoption of smartphones and mobile gaming. An increasing number of systematic reviews have examined the effects of diabetes-related games; however, findings remain inconclusive [9,18-20]. While most interventions have not demonstrated significant clinical effects, their potential educational value, both for patients and the research community, is increasingly acknowledged.

In this study, SugarVita was evaluated in adults with T2DM using a randomized controlled design. Over the 8-week intervention period, participants in the intervention group showed a significant within-group reduction in HbA_{1c} and numerically higher diabetes knowledge scores. While changes in self-confidence and self-management were not statistically significant, outcome measures generally changed in a favorable direction. However, no statistically significant between-group differences were observed after correction for multiple primary outcomes. The relatively small sample size and short intervention duration may have limited the ability to detect statistically significant between-group effects. Participants in both study arms received regular follow-up and structured lifestyle counseling in accordance with national guidelines. SugarVita was therefore evaluated as a complementary educational intervention alongside standard diabetes care, which may partly explain the limited between-group differences observed. In health care systems where access to diabetes education is more limited, follow-up is less frequent, or structured self-management support is less readily available, digital educational tools may yield larger relative effects.

Given these factors, the findings suggest that SugarVita may have contributed to favorable within-group changes in the setting of our study, where participants received instructions and guidance from the research team. Notably, participants rated the game as both enjoyable and educational, which may have contributed to the favorable trends observed in several outcomes. On average, they played for just over 1 hour per week. Higher

XSL•FO
RenderX

total playtime was moderately associated with greater HbA_{1c} reduction, whereas the association with self-management outcomes was weak. Longer or sustained engagement with the game, such as over a longer intervention period, may be required to better evaluate its potential impact. Although these associations were not statistically significant, they may reflect a possible association between engagement and study outcomes.

The findings of this study are consistent with previous research supporting the potential of serious games to enhance diabetes self-management. Early work in children with type 1 diabetes showed that educational games could improve disease-related knowledge and behavior [17]. Among adults with T2DM, game-based interventions have demonstrated varying effects: an autonomous exercise game by Kempf and Martin [18] improved glycemic control and quality of life, and digital health games by Höchsmann et al [19] have shown promising results in promoting physical activity and adherence. Notably, a team-based online game led to significant improvements in HbA_{1c} among US veterans with T2DM [20]. In comparison with previous studies, SugarVita offered a simple, mobile-based game experience that included short educational prompts throughout gameplay. This combination of accessibility, repetition, and engagement may have contributed to the favorable trends observed in diabetes knowledge scores and HbA_{1c} levels.

Several factors may help explain the observed findings associated with SugarVita. The game was co-developed with patients with T2DM, and their preferences, including the types of games they enjoyed, were carefully considered during development. This likely contributed to the game's relevance and acceptability. SugarVita is designed to be easily accessible and simple to integrate into daily life. Players engage in everyday scenarios that reflect real-life situations, offering a sense of recognition and familiarity. Importantly, players receive immediate feedback on how common daily choices influence blood glucose levels, which may enhance learning through direct cause-and-effect visualization. Educational prompts are short and integrated into gameplay, supporting incidental learning—especially for those who may not benefit as much from traditional educational formats such as one-on-one consultations with health care providers, which can sometimes feel abstract or fail to help patients visualize the consequences of their actions. Participant evaluations confirmed that the game was perceived as educational, user-friendly, and especially useful for newly diagnosed patients with diabetes. However, ratings for playability and challenge were moderate, and willingness to continue playing was relatively low, with variation between participants, suggesting that long-term engagement could be improved. Future iterations of SugarVita could address this by introducing more progressive challenges and greater variation in difficulty over time. In addition, expanding social and multiplayer features, such as collaborative or competitive challenges, and improving personalization and visual feedback may help sustain player interest, ensure the game remains both valuable and challenging for returning players, and support longer-term intervention effects. Player performance within the game was not included in the analyses, as comparing in-game scores was complicated by the use of weekly assignments and variation in game modes; therefore, total playtime was used as the main indicator of engagement.

This study has several strengths. It was conducted as a randomized controlled trial, applying a structured design to evaluate the impact of a digital intervention in people with T2DM. Validated instruments were used to assess self-management and self-confidence, while diabetes-related knowledge was assessed using a translated version of the Diabetes Knowledge Test that had not yet been formally validated in Dutch. Real-world HbA_{1c} values were obtained from routine care without requiring additional blood draws, minimizing participant burden. The game was co-designed with the target population, increasing its practical relevance and user engagement. In addition, player data, such as frequency and duration of gameplay, were objectively monitored, enabling the exploration of potential dose-response relationships.

Several limitations should also be noted. The sample size was modest, limiting statistical power, particularly for secondary outcomes. This study was designed as a randomized controlled trial with a parallel-group design; however, as an exploratory pilot study, no formal power calculation was performed. Randomization was stratified based on age and baseline HbA_{1c}; however, some baseline differences remained between groups, particularly in BMI, diabetes duration, and glucose-lowering medication use. Although these differences were not statistically significant, they may have influenced the observed outcomes, especially HbA_{1c}. Baseline use of GLP-1 receptor agonists was more common in the intervention group. Review of medication changes showed both treatment intensifications and discontinuations of GLP-1 receptor agonists within the intervention group, while initiation and dose escalation also occurred in the control group. Nevertheless, the potential influence of baseline differences in glucose-lowering medication use and medication changes during follow-up on glycemic outcomes cannot be excluded. In addition, participants were intentionally recruited on the basis of elevated HbA_{1c} levels (>64 mmol/mol), which may have increased the likelihood of regression to the mean. Therefore, the within-group HbA_{1c} reduction observed in the intervention group should be interpreted cautiously and cannot be attributed solely to the SugarVita intervention.

The diabetes knowledge questionnaire, although based on a widely used instrument, had not been formally validated in Dutch at the time of the study. Consequently, findings related to diabetes knowledge should be interpreted with caution and regarded as exploratory. Future studies should preferably use a formally validated Dutch instrument or validate the translated version prior to implementation. In addition, repeated administration of the same questionnaires may have introduced practice effects, particularly for the diabetes knowledge assessment.

Participant evaluations may have been influenced by social desirability bias and the Hawthorne effect, potentially contributing to more favorable ratings. Furthermore, digital literacy, access to digital devices, and prior experience with mobile apps or games were not formally assessed. These factors

XSL•FO
RenderX

may have influenced engagement with the intervention and its perceived usability.

Furthermore, follow-up was limited to the short-term effects directly following the intervention, so the durability of these benefits remains unknown. In addition, the game was tested in a single center, which may limit the generalizability of the findings to broader or more diverse populations. Finally, the relatively high level of baseline diabetes care in this setting may limit the generalizability of the findings to health care systems with different levels of access to diabetes education.

Conclusions

SugarVita, a mobile serious game co-developed with patients, was perceived as educational and user-friendly by participants. No statistically significant between-group differences were observed for the primary and secondary outcomes. These findings support the feasibility and acceptability of SugarVita as a digital educational intervention for people with T2DM.

Given the global rise in T2DM and the growing need for scalable educational interventions, serious games such as SugarVita may represent a valuable complementary approach to diabetes education and self-management. Accessibility was a key consideration in the game's development. Future work should focus on expanding in-game options and tailoring content to sustain engagement and increase relevance across diverse patient populations. Larger, adequately powered studies with longer follow-up are needed to determine effectiveness and optimize sustained engagement.

Funding

This research was conducted as part of the Diagame project, which was funded by NWO Data2Person (grant 628.011.027).

Authors' Contributions

HH, PW-vP, and ER conceived the study and designed the research. ER and EC-VdM recruited and supervised participants throughout the study period. ER performed the statistical analyses with support from JD and drafted the manuscript. HH, PW-vP, JD, PVG, NvR, and RDV critically revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

HH was a shareholder and served on the board of directors of HRH Diabetes Games B.V. at the time this study was conducted. HH did not have access to the study data and was not involved in the statistical analyses. His contribution was limited to reviewing and commenting on the manuscript as a coauthor. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Assignment card used during the SugarVita intervention.
[PDF File, 622 KB - diabetes_v11i1e99345_app1.pdf]

Multimedia Appendix 2
Participant evaluation form completed after the intervention.
[PDF File, 482 KB - diabetes_v11i1e99345_app2.pdf]

Multimedia Appendix 3
Questionnaires used to assess diabetes knowledge, self-efficacy, and self-management.
[PDF File, 1012 KB - diabetes_v11i1e99345_app3.pdf]

Checklist 1
CONSORT checklist.
[PDF File, 470 KB - diabetes_v11i1e99345_app4.pdf]

References

1. IDF Diabetes Atlas, 10th edition: International Diabetes Federation; 2021.
2. Magliano DJ, Sacre JW, Harding JL, Gregg EW, Zimmet PZ, Shaw JE. Young-onset type 2 diabetes mellitus - implications for morbidity and mortality. Nat Rev Endocrinol 2020 Jun;16(6):321-331. [doi: 10.1038/s41574-020-0334-z] [Medline: 32203408]
3. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev 2013 Jan;93(1):137-188. [doi: 10.1152/physrev.00045.2011] [Medline: 23303908]
4. Tancredi M, Rosengren A, Svensson AM, et al. Excess mortality among persons with type 2 diabetes. N Engl J Med 2015 Oct 29;373(18):1720-1732. [doi: 10.1056/NEJMoa1504347] [Medline: 26510021]

XSL•FO
RenderX

5. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2022 Nov 1;45(11):2753-2786. [doi: 10.2337/dci22-0034] [Medline: 36148880]
6. Franz MJ, MacLeod J, Evert A, et al. Academy of Nutrition and Dietetics nutrition practice guideline for type 1 and type 2 diabetes in adults: systematic review of evidence for medical nutrition therapy effectiveness and recommendations for integration into the nutrition care process. J Acad Nutr Diet 2017 Oct;117(10):1659-1679. [doi: 10.1016/j.jand.2017.03.022] [Medline: 28533169]
7. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-412. [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
8. American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes-2024. Diabetes Care 2024 Jan 1;47(Suppl 1):S20-S42. [doi: 10.2337/dc24-S002] [Medline: 38078589]
9. Reinders EF, de Vries R, Wouters-van Poppel PC, Van Riel NA, Haak HR. Serious digital games for diabetes mellitus: a scoping review of its current state, accessibility, and functionality for patients and healthcare providers. Diabetes Res Clin Pract 2024 Oct;216:111833. [doi: 10.1016/j.diabres.2024.111833] [Medline: 39179114]
10. Maas AH. Playing with numbers: the development of an educational diabetes game [PhD thesis]. : Technische Universiteit Eindhoven; 2017 URL: https://research.tue.nl/en/publications/playing-with-numbers-the-development-of-an-educational-diabetes-g/ [accessed 2026-07-27]
11. Maas AH, van der Molen P, van de Vijver R, et al. Concept development of the Eindhoven Diabetes Education Simulator project. Games Health J 2016 Apr;5(2):120-127. [doi: 10.1089/g4h.2015.0037] [Medline: 26871654]
12. Type 2 diabetes mellitus [Article in Dutch]. Nederlands Huisartsen Genootschap. URL: https://richtlijnen.nhg.org/standaarden/diabetes-mellitus-type-2 [accessed 2026-07-27]
13. Van Der Ven NC, Weinger K, Yi J, et al. The Confidence in Diabetes Self-Care scale: psychometric properties of a new measure of diabetes-specific self-efficacy in Dutch and US patients with type 1 diabetes. Diabetes Care 2003 Mar;26(3):713-718. [doi: 10.2337/diacare.26.3.713] [Medline: 12610027]
14. Bijl JV, Poelgeest-Eeltink AV, Shortridge-Baggett L. The psychometric properties of the Diabetes Management Self-Efficacy Scale for patients with type 2 diabetes mellitus. J Adv Nurs 1999 Aug;30(2):352-359. [doi: 10.1046/j.1365-2648.1999.01077.x] [Medline: 10457237]
15. Fitzgerald JT, Funnell MM, Anderson RM, Nwankwo R, Stansfield RB, Piatt GA. Validation of the revised Brief Diabetes Knowledge Test (DKT2). Diabetes Educ 2016 Apr;42(2):178-187. [doi: 10.1177/0145721715624968] [Medline: 26769757]
16. Maas AH, Rozendaal YJ, van Pul C, et al. A physiology-based model describing heterogeneity in glucose metabolism: the core of the Eindhoven Diabetes Education Simulator (E-DES). J Diabetes Sci Technol 2015 Mar;9(2):282-292. [doi: 10.1177/1932296814562607] [Medline: 25526760]
17. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22(1):77-89. [doi: 10.3109/14639239709089835] [Medline: 9183781]
18. Kempf K, Martin S. Autonomous exercise game use improves metabolic control and quality of life in type 2 diabetes patients - a randomized controlled trial. BMC Endocr Disord 2013 Dec 10;13:57. [doi: 10.1186/1472-6823-13-57] [Medline: 24321337]
19. Höchsmann C, Müller O, Ambühl M, et al. Novel smartphone game improves physical activity behavior in type 2 diabetes. Am J Prev Med 2019 Jul;57(1):41-50. [doi: 10.1016/j.amepre.2019.02.017] [Medline: 31128953]
20. Kerfoot BP, Gagnon DR, McMahon GT, Orlander JD, Kurgansky KE, Conlin PR. A team-based online game improves blood glucose control in veterans with type 2 diabetes: a randomized controlled trial. Diabetes Care 2017 Sep;40(9):1218-1225. [doi: 10.2337/dc17-0310] [Medline: 28790131]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GLP-1: glucagon-like peptide-1
HbA_{1c}: hemoglobin A_{1c}
SGLT2: sodium-glucose co-transporter-2
T2DM: type 2 diabetes mellitus

XSL•FO
RenderX

Edited by GGR Sng, S Li; submitted 24.Apr.2026; peer-reviewed by A Agapiou, Q Fan; revised version received 10.Jul.2026; accepted 11.Jul.2026; published 07.Aug.2026.

Please cite as:
Reinders E, Cruts - Van de Meulengraaf E, De Vries R, van Riel N, Van Gorp P, Dieleman J, Wouters - van Poppel P, Haak H
A Serious Digital Game (SugarVita) to Support Diabetes Self-Management: Pilot Randomized Controlled Trial
JMIR Diabetes 2026;11:e99345
URL: https://diabetes.jmir.org/2026/1/e99345
doi:10.2196/99345

© Edouard Reinders, Elma Cruts - Van de Meulengraaf, Ryan De Vries, Natal van Riel, Pieter Van Gorp, Jeanne Dieleman, Pleun Wouters - van Poppel, Harm Haak. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 7.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Integration of Consumer Activity Trackers Into a Community Weight Management Intervention to Support Physical Activity in Adults at Risk for or With Type 2 Diabetes: Mixed Methods Study Using the RE-AIM Framework

William Hodgson[1*], BSc, MRES, PhD; Alison Kirk[1*], BSc, PhD; Marilyn Lennon[2*], BSc, PhD; Xanne Janssen[1*], BSc, PhD; David Kennedy[1*], BSc

[1]Department of Psychological Sciences and Health, Physical Activity for Health Research Group, University of Strathclyde, 40 George Street, Glasgow, Scotland, United Kingdom

[2]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, Scotland, United Kingdom

[*]all authors contributed equally

Corresponding Author:
William Hodgson, BSc, MRES, PhD
Department of Psychological Sciences and Health, Physical Activity for Health Research Group, University of Strathclyde, 40 George Street, Glasgow, Scotland, United Kingdom

Abstract

Background: Type 2 diabetes affects 483 million adults worldwide, with rising prevalence and an estimated 6 million premature deaths annually. Low physical activity is a key risk factor, while increased activity can reduce disease onset and improve metabolic health. Consumer activity trackers, when paired with behavior change strategies, have shown potential to increase physical activity among adults with type 2 diabetes.

Objective: This study explored the integration of consumer activity trackers into a community-based weight management intervention to support physical activity in adults at risk for or living with type 2 diabetes.

Methods: A mixed methods design was used to generate a comprehensive understanding of implementation. Participants were recruited during registration for "Weigh to Go," a community-based weight management program in Lanarkshire, Scotland. Health care professionals delivering the intervention were recruited by email. Participants received a Fitbit Charge 5 to monitor daily steps and moderate-to-vigorous–intensity physical activity. Semistructured interviews were conducted with 10 participants and 10 health care professionals. Qualitative data were analyzed thematically, and quantitative analysis examined changes in recorded Fitbit activity data. Findings were interpreted using the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework.

Results: Daily steps increased significantly between week 1 and week 7 (mean difference 5345 steps; P=.002). Qualitative findings highlighted 5 themes. First, providing devices free of charge enhanced reach by removing financial barriers. Second, educational classes were considered essential for effectiveness, particularly instruction on device use and interpretation of activity data. Third, staff expressed a need for greater understanding of device functionality and data outputs, supporting broader adoption of trackers within weight management services. Fourth, managers would benefit from a detailed protocol outlining tracker introduction, use, data analysis procedures, evaluation metrics, and costs to ensure efficient and consistent implementation. Fifth, extending compulsory attendance at intervention sessions was considered important for long-term maintenance of behavior change. The observed decline in moderate-to-vigorous–intensity physical activity after week 7 was attributed to challenges in sustaining engagement beyond the structured phase of the program.

Conclusions: This study demonstrates the feasibility of integrating consumer activity trackers into a community-based weight management intervention. Applying the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework revealed that free device provision, participant and staff education, clearly defined implementation protocols, and structured attendance expectations can strengthen tracker-supported interventions. The use of the Fitbit device as both a measurement and intervention tool also raises methodological considerations, emphasizing the need for future research to differentiate these dual roles. Overall, activity trackers show promise for supporting physical activity among adults at risk for or with type 2 diabetes when embedded within well-designed community programs. These findings underscore the importance of aligning technological tools with supportive behavioral strategies to maximize health outcomes. The results also highlight opportunities for refining community programs through closer integration of digital health tools.

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e91073) doi:10.2196/91073

KEYWORDS

type 2 diabetes; physical activity; sedentary behavior; Fitbit; activity tracker; adults; Weigh to Go; weight management; RE-AIM framework; Reach, Effectiveness, Adoption, Implementation, and Maintenance

Introduction

Type 2 diabetes mellitus is a chronic noncommunicable disease that occurs when the body cannot produce enough insulin, which facilitates the uptake of glucose into cells. This, in turn, raises blood glucose levels. If left untreated, type 2 diabetes can lead to premature death. Complications of type 2 diabetes include a higher risk of developing cardiovascular disease, retinopathy, neuropathy, and nephropathy [1]. Globally, it is estimated that 483 million adults (aged 20 - 79 years) are living with type 2 diabetes, and by 2045 this number is expected to rise to 700 million [2]. In the United Kingdom, 5 million adults have been diagnosed with type 2 diabetes, and their treatment costs the National Health Service £12 billion each year (or US $16.8 billion each year; conversion rate: £1=US $1.34) [3]. Major risk factors for developing type 2 diabetes are physical inactivity, sedentary behavior, and being overweight or obese [1]. Adults diagnosed with type 2 diabetes have been shown to be less physically active and spend more time being sedentary compared with those who do not have the disease [4]. The American Diabetes Association has stated that physical activity can help prevent adults from developing type 2 diabetes, and for those already diagnosed with the disease, it can improve their diabetes and general health [5]. Even low-intensity physical activity and a reduction in sedentary behavior can bring health benefits for adults diagnosed with type 2 diabetes by lowering their blood glucose levels and BMI [6,7].

Physical activity interventions can significantly reduce the blood glucose levels of adults diagnosed with type 2 diabetes [6-8]. Physical activity interventions that incorporate wearable technology, such as consumer activity trackers and pedometers, have been developed for adults diagnosed with chronic noncommunicable diseases, including those living with type 2 diabetes [9]. Consumer activity trackers, including Fitbit models, are valid and reliable devices for measuring users' daily steps and physical activity intensity [10]. Physical activity interventions that incorporate a consumer activity tracker have been shown to lower blood glucose levels, reduce BMI, increase activity, and reduce sedentary time in adults living with type 2 diabetes [11]. Physical activity interventions that combine an activity tracker with either a behavioral change intervention or a web-based platform can bring significant health improvements for patients with type 2 diabetes [12]. In addition, consumer activity trackers have the potential to allow health care professionals to remotely monitor individuals' physical activity as part of their clinical care and for the person to self-manage their activity through feedback from the device [13]. Community-based weight management interventions are 1 way to try and support the use of activity trackers by overweight adults and promote physical activity in this population [14]. The Weigh to Go program is an established community- based weight management intervention that aims to support adults to increase physical activity and improve diet. However, referral and adherence of adults with type 2 diabetes to the program have been low, and there has been limited evaluation of longer-term physical activity behavior change [14]. Integrating consumer activity trackers into such programs may help address these gaps by enhancing engagement, supporting self-monitoring, and providing continuous feedback. A further methodological consideration is that the activity tracker functions simultaneously as a measurement tool and an intervention component. This dual role can influence participant behavior through feedback, goal setting, and accountability features, meaning that the act of measurement may itself contribute to behavior change [13]. Recognizing this overlap is important for interpreting outcomes and for designing future studies that aim to disentangle these effects.

This study aimed to explore, through a mixed methods study using the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework, the integration of consumer activity trackers into a community weight management intervention to support physical activity in adults at risk for or with type 2 diabetes. The RE-AIM framework was used not only as an evaluative structure but also as a sensitizing guide for data collection, enabling exploration of how the intervention reached participants, how it was experienced, how staff adopted and delivered it, and how it might be sustained in routine practice.

Methods

The Study Context: Weigh to Go Intervention

The Weigh to Go intervention is an established community-based weight management behavior change program based in Lanarkshire, Scotland. Sessions were held within 12 local community centers and sports centers. The intervention consists of a 15-week active phase and a 15-week maintenance phase. In the active phase, participants receive a 45-minute educational workshop and a 45-minute circuit-based physical activity session each week. Each session has been developed to stand alone, and evidence-based content has been developed by dietitians, nurses, and psychologists [14]. Sessions include discussion and reflection on dietary education content, as well as activities to support positive and sustainable behavior change through motivation-, action-, and prompt-informed health behavior change strategies [15]. In the maintenance phase, participants receive a short session with reflection on what was covered during the equivalent week in the active phase, alongside 45 minutes of physical activity. Attendance at the maintenance sessions is optional, whereas attendance at the active phase classes is expected. The standard physical activity component is circuit-based, with adaptations to accommodate most levels of fitness and mobility. The intervention is free for participants and is delivered in local leisure centers and community facilities. The Weigh to Go intervention has the

XSL•FO
RenderX

potential to be an effective community-based program to support weight management and physical activity for adults with type 2 diabetes and those at high risk of developing the disease. However, referral and adherence of adults with type 2 diabetes to the Weigh to Go intervention have been low, and there has been limited evaluation, particularly of longer-term physical activity behavior change. Health care professionals also report limited follow-up on the progress of people they have referred [14].

Justification for the Addition of Activity Trackers Into the Weigh to Go Program

In weight management programs such as Weigh to Go, client referral and sustained adherence are crucial determinants of success. Yet, traditional interventions relying primarily on dietary advice and tracking, with limited assessment of physical activity, often face high dropout rates [16]. The integration of wearable activity trackers could address these gaps and support self-monitoring of physical activity, behavior change, and ongoing engagement. Patients referred to weight management services frequently disengage early; lack of ongoing feedback contributes to this attrition [17]. Wearables provide continuous self-monitoring, one of the most effective behavior change techniques, which has been shown to mediate weight loss. Moreover, users who consistently track physical activity lose significantly more weight than those with low adherence [18].

Wearables deliver critical behavior change strategies. Self-monitoring and goal setting can be achieved through step tracking and recording of active minutes, which enhances motivation and self-awareness [19]. Feedback and habit formation can be achieved through automated reminders to support long-term adherence [20]. Providing professional oversight through consultations has been shown to significantly boost physical activity outcomes [21]. Meta-analyses show that wearables improve activity levels (~1800 additional steps/d, 40 minutes walking/d) and reduce body weight (~1 kg) across clinical populations [17]. In overweight adults, activity tracker–inclusive interventions outperformed standard programs over ≤6 months, particularly among middle-aged and older adults [16]. Another meta-analysis reported moderate-to-large standardized increases in daily steps (standardized mean difference 0.54) and moderate-to-vigorous physical activity (standardized mean difference 0.47), with meaningful reductions in weight and BMI [22].

The Weigh to Go program typically secures referrals based on clinical or motivational criteria. However, post referral, clients often lose momentum [14]. Embedding wearables at the referral point could potentially increase immediate engagement by giving clients tangible metrics and daily goals and produce early positive results for participants. This could also encourage long-term adherence and provide continuous data to tailor individual support. Pairing trackers with professional guidance has the potential to create a multimodal intervention—diet, activity, technological feedback, and counseling—which yields better short-term outcomes than standard programs alone [16].

Incorporating physical activity trackers into the Weigh to Go program has the potential to address critical gaps in referral and adherence. They operationalize evidence-based behavior strategies, including self-monitoring, goal setting, and feedback, supported by robust meta-analytic evidence. When paired with professional oversight, trackers enhance engagement, mitigate dropout, and drive measurable improvements in activity and weight loss. This technology-enhanced model not only aligns with best practices in behavioral medicine but also leverages scalable tools to boost program effectiveness and sustainability.

Participants

Adults Registered on the Community Weigh to Go Intervention

Participants were recruited via a study flyer at various venues hosting the registration stage of the Weigh to Go intervention. The initial inclusion criteria included adults aged >18 years (no upper limit set), diagnosed with type 2 diabetes or at risk of developing the disease, residing in the United Kingdom, having access to internet services, and being able to read and write in English. Exclusion criteria included not being diagnosed with type 2 diabetes or not being at risk of developing the disease, or having been advised by a health care professional not to undertake physical activity. Participants interested in taking part in the study were asked to email the research team. The study participant information sheet and consent form were uploaded onto the secure Qualtrics (Silver Lake) survey system. A link to this form was emailed to interested participants, and their consent was recorded by indicating “yes” on the Qualtrics consent form. Once consent was received, participants were emailed a link to a baseline questionnaire on the Qualtrics system. This questionnaire gathered demographic details, educational level, current activity tracker use, and health conditions.

Health Care Professionals

Health care professionals (n=10) involved in the delivery and onboarding of the Weigh to Go intervention were recruited via an email from the research team. Recruitment criteria included adults aged >18 years, residing in the United Kingdom, and being a health care professional (eg, doctor, nurse, fitness instructor, or Weigh to Go intervention administrator). The study participant information sheet and consent form were uploaded onto the secure Qualtrics survey system. A link to this form was emailed to participants, and their consent was recorded. Once consent was received, a link to a Qualtrics questionnaire was emailed to the participant to gather information regarding demographics, qualifications, and job role characteristics.

Procedure

A summary of the study procedures for the Weigh to Go intervention participants is provided in Figure 1. A summary of the study procedures for the Weigh to Go health care professionals is provided in Figure 2.

XSL•FO
RenderX

Figure 1. Flow diagram showing summary of the study procedures for the Weigh to Go intervention participants.

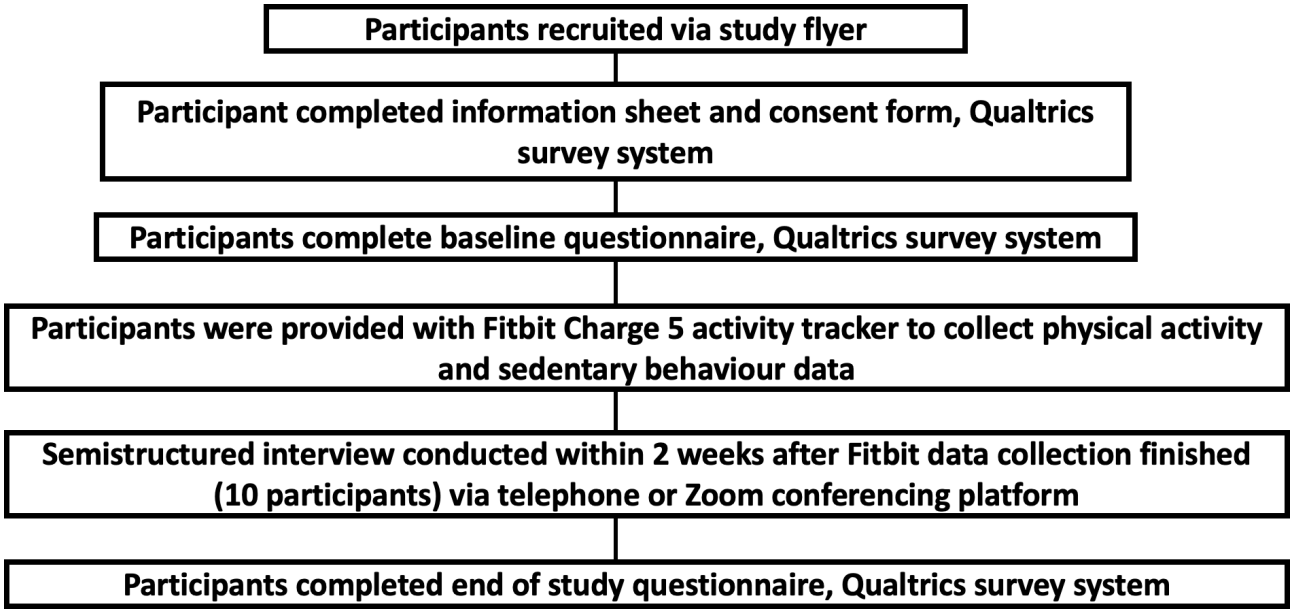
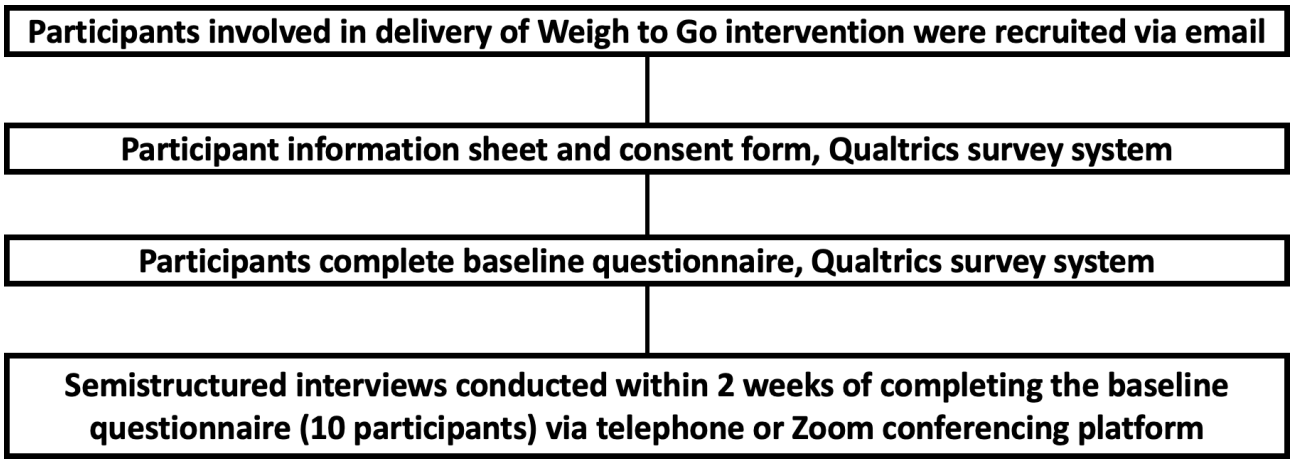


Figure 2. Flow diagram showing a summary of the study procedures for the Weigh to Go intervention for health care professionals.



Data Collection

Quantitative Data Collection

Once participants had consented and completed the baseline survey, they were provided with a Fitbit Charge 5 device (provided free of charge by Fitbit United Kingdom), which they were allowed to keep at the conclusion of the study. The research team set up an individual Fitbit account for each participant (Google Gmail email address and password) to allow both the participant and study team to access the activity data from the Fitbit device. Instructions on how to access this account and set up the Fitbit device were emailed to each participant. Additional support was provided, if required, via email, telephone, or face-to-face contact with the participants at the Weigh to Go venue. Participants were asked to use the Fitbit device over the period during which they attended the Weigh to Go intervention. Once each participant had finished attending the Weigh to Go intervention, they were sent an email containing a link to the Qualtrics end-of-study questionnaire. The research team downloaded the participants' Fitbit activity data from the Fitbit account. The downloaded data included daily steps and weekly minutes of moderate-to-vigorous–intensity physical activity. The Fitbit data and the baseline and end-of-study questionnaire data were uploaded into SPSS (IBM Corp) statistical software for analysis.

Qualitative Data Collection

This study adopted a qualitative design to explore experiences of Fitbit use and engagement with the Weigh to Go intervention. The methodological orientation was constructivist and interpretivist, recognizing that knowledge is coconstructed through interaction and shaped by social and contextual factors [23,24]. Data were generated through semistructured interviews and analyzed using reflexive thematic analysis [25-27]. Reporting followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) 32-item checklist, ensuring transparency across domains of research team and reflexivity, study design, and analysis [28].

A semistructured interview schedule was designed around the 5 RE-AIM dimensions [29]. These dimensions provided a structured yet flexible scaffold for exploring experiences of intervention use and delivery.

XSL•FO
RenderX

An abductive interviewing approach was adopted [30], enabling the researcher to attend to participants' lived experiences while actively probing surprising, contradictory, or unanticipated insights. This iterative, dialogical style supported the generation of richer explanations by moving fluidly between empirical data and sensitizing theoretical concepts [31].

Interviews were conducted via secure Zoom (Zoom Video Communications) video conferencing or by telephone, based on participant preference. Each interview lasted between 45 and 90 minutes, was audio-recorded on an encrypted digital recorder, and then transcribed verbatim. Transcripts were anonymized, checked for accuracy, and uploaded to NVivo (version 12; Lumivero) for data management.

Study Design and Participants

One-to-one interviews were conducted with two groups: (1) ten Weigh to Go participants purposively sampled from those who provided Fitbit data, and (2) ten health care professionals drawn from the staff members involved in delivering the intervention.

Although participants were randomly selected within these eligible pools, the design was guided by a purposive sampling logic, ensuring recruitment of individuals directly engaged with Fitbit use or intervention delivery [24]. This approach balanced fairness in selection with the need for information-rich accounts. Interview questions were mapped to the RE- AIM dimensions. For example, "reach" questions explored barriers to joining or staying in the program; "adoption" questions asked staff about organizational readiness; "implementation" questions examined how trackers were introduced and supported; and "maintenance" questions invited participants and staff to reflect on whether and how tracker use could continue beyond the active phase. These questions focused on participants' experiences and expectations rather than predictions of future behavior.

Sample size in qualitative research is not determined by statistical representativeness but by the adequacy of data to address research questions and generate meaningful insights [32,33]. The decision to recruit 20 participants in total was informed by methodological principles of information power, which suggest that smaller samples may be sufficient when study aims are specific, participants are closely aligned with the topic of inquiry, and the analysis is detailed [32]. Reflexive thematic analysis does not prescribe a fixed number of interviews; rather, it emphasizes depth, variation, and interpretive engagement over numeric thresholds [26]. Empirical benchmarks indicate that major thematic patterns can often be captured with 12 - 15 interviews in relatively homogeneous samples [34,35], although professional perspectives typically add diversity and nuance [36]. Accordingly, this study adopted a pragmatic balance between analytic richness and feasibility, monitoring sufficiency throughout analysis and justifying the final sample size on grounds of analytic adequacy rather than data saturation as a fixed endpoint [26,33].

Justification for Abductive Analysis

The study used abductive analysis because neither purely inductive nor deductive logic was sufficient for the research aims. Inductive approaches, while useful for grounded, bottom-up coding, risk limiting interpretation to surface-level description without connecting findings to broader explanatory frames [24]. Deductive approaches, by contrast, can overconstrain interpretation through rigid application of preexisting theories, missing novel, or contextually embedded meanings [37].

Abduction provided a middle ground, an iterative process of moving between empirical observations and theoretical frameworks [38]. In this study, abduction was particularly well suited because (1) surprise and anomaly were expected. Weigh to Go participants' and health care professionals' experiences with Fitbit use could not be fully anticipated, and abduction-supported theorizing around unexpected findings; (2) the RE-AIM framework informed but did not restrict analysis. Abduction enabled findings to dialog with RE-AIM concepts without being forced into them; and (3) the goal was to generate plausible explanatory insights. Abduction encourages researchers to adjudicate between rival explanations, ensuring that final themes represent the most coherent and plausible interpretations [31].

Thus, abduction was methodologically congruent with reflexive thematic analysis, which views researcher subjectivity and theoretical engagement as analytic resources rather than as sources of bias [26].

Data Analysis

The Framework to Guide Analysis

The RE-AIM framework is a planning and evaluation tool used to improve the adoption and sustainable implementation of health-related interventions from research settings into real-world working environments [29]. The main RE-AIM dimensions are RE-AIM. Reach is defined as the absolute number, proportion, and representativeness of individuals participating in each initiative. Effectiveness is the impact of an intervention on outcomes, including potential negative effects, quality of life, and health components. Adoption is the proportion, representativeness, and absolute number of organizational agents involved in the intervention. Implementation is considered at an organizational level and refers to how the program was delivered by staff. Maintenance is defined as the extent to which a program or policy becomes embedded in routine practice [29]. At an individual level, maintenance is a measure of the long-term impact of an intervention over 6 months [39].

Quantitative Data Analysis

Descriptive analysis and repeated measures 1-way ANOVA tests were undertaken on the Fitbit data collected. One-way within-subjects ANOVA statistical tests were conducted to measure the effect of the participants' use of the Fitbit Charge 5 activity tracker on their physical activity over 20 weeks. Magnitude of change was measured between week 1, week 7, week 15, and week 20 in relation to daily steps taken (number of steps) and weekly moderate-to-vigorous–intensity physical activity (minutes). The dependent variables were (1) mean daily step count and (2) mean weekly minutes of moderate-to-vigorous–intensity physical activity (MVPA). The independent variable was time, with 4 repeated measures corresponding to week 1, week 7, week 15, and week 20. This

XSL•FO
RenderX

design enabled assessment of within-subject changes in physical activity across the intervention and immediate postintervention phases. Although the repeated-measures ANOVA assumes normally distributed data, inspection of residuals indicated that deviations from normality were minimal and within acceptable bounds for parametric testing. Given the moderate sample size, the ANOVA was considered robust to minor violations of normality. However, a nonparametric alternative (Friedman test) was also considered. In exploratory analyses, results from the Friedman test were consistent with the ANOVA findings, supporting the validity of the conclusions. Nonetheless, the parametric approach was retained as the primary method due to its greater interpretability and the ability to report effect sizes. The choice of time points (week 1, week 7, week 15, and week 20) was methodologically and practically driven. Week 1 acted as the initial reference point at the start of the intervention. Week 7 corresponded with the midpoint of the 15-week active phase, allowing assessment of early change and adherence. Week 15 marked the end of the active intervention phase, providing a natural point for evaluating intervention effects. Week 20 was included to assess whether physical activity behavior was maintained after the active phase concluded, aligning with the RE-AIM framework's emphasis on maintenance. Data beyond week 20 (ie, weeks 25 and 30) were too sparse for valid statistical analysis and could have risked compromising participant anonymity.

Qualitative Data Analysis

Data analysis followed Braun and Clarke's reflexive thematic analysis. NVivo 12 was used for organizing codes and data extracts, but interpretation remained the authors' responsibility. Analysis was iterative and abductive, proceeding through Braun and Clarke's six phases:

1. Familiarization: the author immersed himself in transcripts and audio, producing reflexive memos that captured first impressions, contradictions, and anomalies.
2. Generating initial codes: coding was conducted across the dataset at both semantic (descriptive) and latent (interpretive) levels. Abductive reasoning supported flexible refinement of codes when unexpected insights emerged.
3. Constructing candidate themes: codes were collated into candidate themes representing shared patterns of meaning relevant to the research question. Rival candidate structures were considered and abductively evaluated for explanatory power.
4. Reviewing and refining themes: candidate themes were tested against data extracts and the dataset as a whole. Incoherent or redundant themes were reworked, merged, or discarded. Negative cases were actively sought to challenge emerging interpretations.
5. Defining and naming themes: each theme was clearly defined with a central organizing concept and explanatory warrant. For example, the theme "Fitbit as accountability partner" captured both motivational and burdensome aspects of device use.
6. Producing the report: themes were written into an analytic narrative, weaving data extracts with interpretation to build plausible explanations that extended beyond description. Quotations were included as analytic evidence, not mere illustrations [26].

Reflexivity and Rigor

In line with COREQ principles [28], reflexivity was integral throughout. Regular supervisory meetings provided reflexive dialog and guarded against premature closure of analysis. Trustworthiness was addressed using strategies congruent with reflexive thematic analysis [40,41]: (1) credibility through prolonged engagement with data, reflexive memoing, and analytic dialog; (2) dependability via a transparent audit trail documenting coding iterations and theme refinements; (3) confirmability by explicitly acknowledging researcher subjectivity as a resource in interpretation; and (4) resonance and contribution by producing findings with conceptual and practical utility for digital health intervention design.

Ethical Considerations

Ethical approval for this mixed methods study was obtained from the University of Strathclyde Ethics Committee (approval no 2454). For the study, both adults diagnosed with type 2 diabetes and health care professionals were provided with web-based participant information sheets. Informed consent was obtained through the digital signing of a consent form. Participants electronically selected individual items in the digital form, corresponding to the paper consent form, to confirm that they had read and agreed to each item. Their electronic signature was achieved by entering their allocated 4-digit identification number. The research team undertook a number of steps to ensure the security of the information collected. To ensure anonymity, each participant was allocated a unique 4-digit identification number, and all data were stored under this number. Any information stored by the research team was stored within university-approved, password-protected, encrypted storage sites. Participants involved in this study received no compensation for taking part.

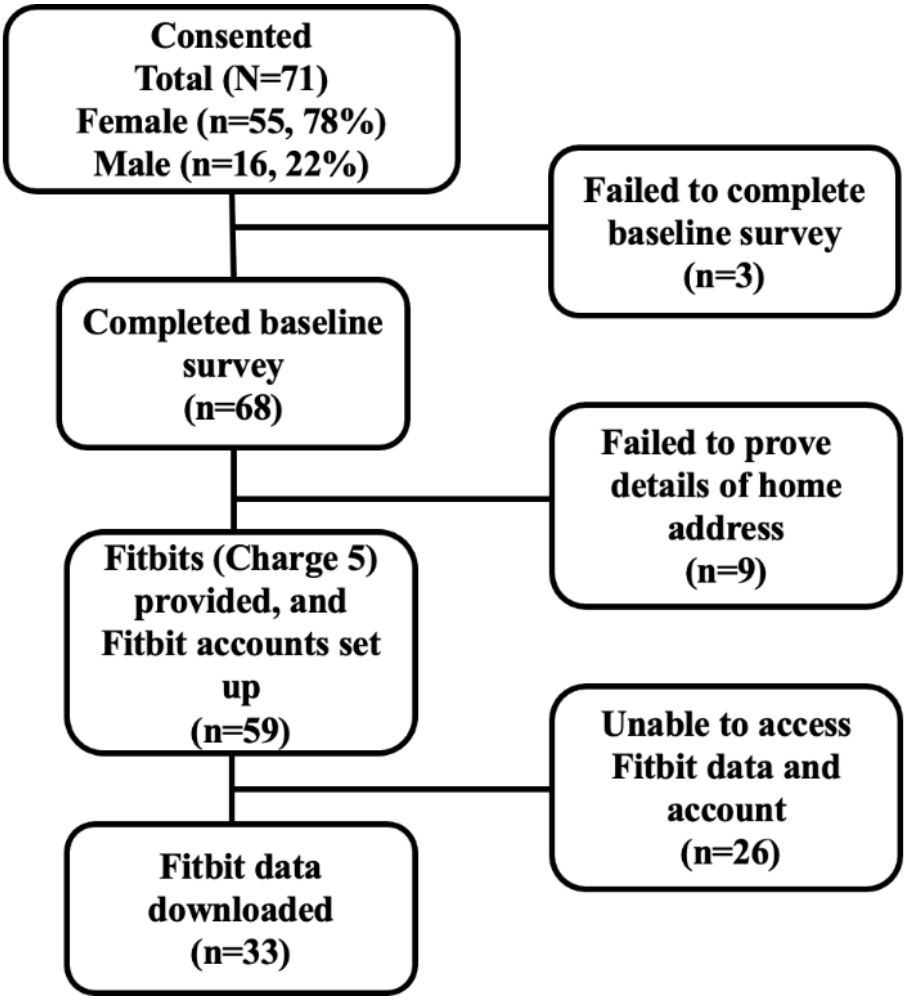
Results

Weigh to Go Participants Recruitment and Retention

Weigh to Go participants were adults who had either volunteered to take part in the program or had been referred by a health care professional. Details of the Weigh to Go participant recruitment and retention are provided in Figure 3.

XSL•FO
RenderX

Figure 3. Weigh to Go participant recruitment and retention.



Explanation of Fitbit Data Loss and Its Implications

In research involving digital health technologies, data integrity is paramount. Wearable activity trackers such as Fitbit devices offer researchers valuable opportunities to collect continuous, real-world data on physical activity. However, technological and procedural challenges can hinder data acquisition, leading to significant data loss. This was observed in this study, where Fitbit data loss affected both the completeness of analyses and the generalizability of findings.

Figure 3 illustrates participant recruitment and retention throughout the Weigh to Go intervention. Although participants were provided with Fitbit devices and accounts were set up by the research team, unexpected changes to Google's security protocols during the study period caused substantial disruption. Specifically, midway through data collection, Google introduced additional 2-step verification processes for Fitbit accounts. This required participants to authorize access to their Fitbit data via a smartphone code linked to their Fitbit mobile app. Consequently, the research team could no longer directly access the accounts to download data.

To mitigate this, the team attempted to schedule telephone authorizations with participants during the download process. However, many participants (n=26) failed to respond to emails or were otherwise unavailable, which prevented access to their data. This issue ultimately resulted in the loss of approximately 26 (44%) participant Fitbit datasets, thereby contributing to the attrition shown in Figure 3. This data loss has implications for interpreting the quantitative findings, particularly regarding representativeness and the ability to assess integration across the full participant cohort.

Participants Demographics

Weigh to Go participant and health care professional demographics are provided in Tables 1 and 2. For the Weigh to Go participants, these included age, gender, educational qualifications, ethnicity, area of residence, health conditions, and previous use of an activity tracker. For the health care professionals, these included age, gender, educational qualifications, and ethnicity.

XSL•FO
RenderX

Table . Weigh to Go participant demographics.

Characteristic	Value
Age (years), mean (SD)	54.6 (12.24)
Sex, n (%)	
Male	6 (18)
Female	27 (82)
Educational qualifications, n (%)	
Degree	9 (27)
SVQ[a]	9 (27)
School qualifications	6 (18)
Higher education	7 (21)
Other qualifications	2 (7)
Ethnicity, n (%)	
White	33 (100)
Area of residence, n (%)	
Rural	3 (9)
Urban	30 (91)
Health conditions, n (%)	
Type 2 diabetes	17 (52)
Arthritis	9 (27)
Mental health	5 (15)
Asthma	3 (9)
COPD[b]	1 (3)
Stroke	1 (3)
Heart disease	1 (3)
Learning difficulties	1 (3)
Previous use of an activity tracker, n (%)	
Yes	1 (6)
No	31 (94)

[a]SVQ: Scottish Vocational Qualification.
[b]COPD: chronic obstructive pulmonary disease.

XSL•FO
RenderX

Table . Health care professional participant demographics.

Characteristic	Value
Age (years), mean (SD)	51.3 (9.7)
Sex, n (%)	
Male	3 (30)
Female	7 (70)
Educational qualifications, n (%)	
Degree	7 (70)
SVQ[a]	3 (30)
Ethnicity, n (%)	
White	9 (90)
Asian	1 (10)

[a]SVQ: Scottish Vocational Qualification.

Quantitative Data Analysis

Results are presented under the appropriate RE-AIM main dimensions.

Reach

Seventy-one adults consented to take part in the study between July 2023 and January 2024. Of those who consented, 55 (78%) were female, and 16 (22%) were male. All 71 participants were White. We were unable to calculate the total number of Weigh to Go participants approached to take part in this study, as this was not recorded by the intervention staff. In total, 705 participants attended the Weigh to Go session venues targeted for this study (n=71, 10% of these consented to take part). The average number of participants attending each venue was 10.

Effectiveness

Weigh to Go Participants' Fitbit Daily Steps (Number Taken)

A test of normality was carried out, and the assumption was met. The Mauchly test of sphericity produced a nonsignificant result (P=.62). Reporting sphericity assumed, there was a significant, large effect of Fitbit use on daily steps ($F_{3,45}$=5.93; P=.002; η^2=0.19). Post hoc comparisons using the Bonferroni correction for multiple comparisons were carried out. There was a significant increase in daily steps between week 1 and week 7 (P=.008), with participants in week 7 walking on average 5345 more steps. There was no statistically significant difference in steps between any of the other weeks.

Weigh to Go Participants Fitbit Weekly Minutes of Moderate to Vigorous Intensity Physical Activity

A test of normality was carried out and the assumption was met. Mauchly test of sphericity produced a significant result (P=.04). Reporting Greenhouse-Geisser correction showed that there was no significant effect of Fitbit use on weekly minutes of moderate-to-vigorous–intensity physical activity ($F_{1.68,13.45}$=0.65; P=.51; η^2=0.07). The decline in MVPA at weeks 15 and 20 reflects the transition from the structured active phase to the optional maintenance phase. Participants reported reduced accountability, fewer scheduled sessions, and diminished novelty of the device, all of which contributed to reduced engagement. This pattern suggests that the integration of trackers was most effective when embedded within structured program components, and less effective when participants were expected to self- manage without ongoing support.

This section evaluates how well the activity tracker was integrated into the program, rather than the program's overall effectiveness. The observed changes in steps and MVPA illustrate how participants engaged with the tracker within the program context, highlighting the conditions under which integration supported behavior change.

Adoption

The total number of staff involved in the delivery of the Weigh to Go intervention is provided in Table 3.

Table . Total number of staff involved in the delivery of the Weigh to Go intervention.

Role	Staff, n (%)
Intervention instructor	22 (55)
Project manager	8 (21)
Administration support	5 (12)
Researchers	5 (12)

XSL•FO
RenderX

The types and number of venues used for the delivery of the Weigh to Go intervention and the total number available throughout the 2 local councils are provided in Table 4.

Table . Types of venues used to deliver the Weigh to Go intervention.

Venue type	Number used, n (%)	Total number available, n
Community center	8 (7)	108
Community fitness center	4 (14)	29
Total	12 (9)	137

The adoption of the activity tracker varied across staff roles and settings. Fitness instructors and frontline staff were generally enthusiastic, viewing the tracker as a motivational tool that complemented existing program components. Administrative staff expressed uncertainty about their role in supporting device use, highlighting gaps in training and clarity.

Organizational adoption was influenced by (1) perceived relevance of trackers to program goals, (2) staff confidence in explaining device features, (3) availability of time during onboarding sessions, and (4) existing digital literacy among staff and participants.

Some staff described the tracker as an “add-on” rather than an integrated component, indicating partial adoption at the organizational level.

Implementation

Signposting for the Weigh to Go intervention used by the study participants is provided in Table 5.

Table . Weigh to Go intervention signposting.

Signposting source	Value, n (%)
Diabetes nurse	5 (15)
Practice nurse	5 (15)
Local gym	5 (15)
GP[a]	3 (9)
Local community center	5 (15)
Website	4 (13)
Work email	1 (2)
Word of mouth	2 (7)
Facebook	2 (7)
Carer center	1 (2)

[a]GP: general practitioner.

The costs of delivering the Weigh to Go intervention included session cost per 15 weeks (US $672), session cost per 30 weeks (US $1344), venue cost per year (US $2016), annual intervention cost (US $60,480), average cost per participant per the 15-week active phase (US $67), and average cost per participant per the 30-week active and maintenance phase (US $134). Though the Fitbit devices were supplied free of charge for this study, the individual retail cost of purchasing the Fitbit Charge 5 was US $136.37. The total retail cost for this study would have been US $8046.03.

The mean number of sessions Weigh to Go participants attended during the 15-week active phase of the Weigh to Go intervention was 11.35 (SD 3.63) sessions.

Fidelity of Delivery

Implementation fidelity varied across sites. Some instructors consistently introduced the tracker during week 1, demonstrated its features, and checked usage during sessions. Others provided minimal explanation due to time constraints or uncertainty about the device. Participants reported inconsistent messaging, with some receiving detailed guidance and others receiving only basic setup instructions.

Barriers to Implementation

Key barriers to implementation included limited staff training on device features, competing demands during onboarding sessions, technical issues (eg, syncing and account access), and variability in participants’ digital literacy.

Facilitators of Implementation

Facilitators of implementation included staff enthusiasm, participants’ curiosity and motivation, clear visual feedback from the device, and integration with weekly physical activity sessions.

These findings indicate that implementation was feasible but uneven, with fidelity dependent on staff confidence and available time.

Maintenance

The mean number of sessions all Weigh to Go participants attended during the maintenance phase of the Weigh to Go

XSL•FO
RenderX

intervention was 1.91 (SD 2.57). Only 1 participant attended the full 15 weeks of the active and maintenance phases of the intervention.

Only 5 Weigh to Go participants provided Fitbit data beyond 20 weeks. The mean number of maintenance sessions attended by these 5 participants was 12.2 (SD 3.47).

Individual-Level Maintenance

Participants described mixed intentions to continue using the tracker after the program. Some planned to maintain daily step goals, others anticipated reduced use without structured support, and several noted that the device became less motivating once novelty faded.

The decline in MVPA at weeks 15 and 20 supports these qualitative accounts, indicating challenges in sustaining engagement without ongoing program structure.

Setting-Level Maintenance

Staff expressed interest in continuing to use trackers but highlighted the need for a formal protocol, clearer guidance on data use, consideration of long-term device costs, and integration into routine monitoring systems.

These findings suggest that maintenance at the organizational level would require structural support, resource allocation, and clearer processes.

Qualitative Findings

Abductive thematic analysis of interviews conducted with the Weigh to Go participants (n=10) and health care professionals (n=10) involved in the delivery and onboarding of the Weigh to Go intervention identified 6 main themes and 11 subthemes. The themes are provided in Table 6 and discussed in more detail below.

Table . Study main themes (n=5) and subthemes (n=11).

Main themes	Subthemes
Fitbit use	• 1.1 Barriers • 1.2 Enablers
Promotion	• 2.1 Health care • 2.2 Weigh to Go intervention
Clinical care	• 3.1 Benefits of activity tracker use • 3.2 Use by health care staff
Finance	• 4.1 Costs (patient and health care provider) • 4.2 Health care monetary savings
Fitbit functions	• 5.1 Preferences • 5.2 Interpretation
Effectiveness	• Physical activity • Goal setting

Fitbit Use

Participants described several barriers that limited their ability or willingness to use activity trackers. A central issue was low motivation to engage in physical activity, which diminished the perceived utility of the device. One participant reflected:

> *I am not interested in exercising even though I know it is good for me. I just don't have the drive to exercise.* [Female participant, 62 years]

Such accounts illustrate that, while Fitbit devices can support behavior change, they are unlikely to generate motivation where this is fundamentally absent.

A further barrier related to digital literacy is particularly prevalent among older adults. As 1 health care professional noted:

> *I think some of the older participants will struggle using activity trackers as they have poorer technology skills compared with the younger generation.* [Female professional, 38 years]

This concern highlights the persistence of a digital divide that risks excluding certain groups unless targeted support is provided.

Financial constraints also emerged as a significant limitation. One professional commented:

> *I know that some of those taking part in the Weigh to Go programme have little money and could not afford a Fitbit and do not own a mobile phone because of the cost.* [Female professional, 47 years]

This illustrates that cost-related barriers extend beyond the device itself to include the ownership of compatible technologies and ongoing data costs.

For others, health conditions imposed practical limitations on activity tracker use. A participant with multiple chronic conditions explained:

> *I have underlying health conditions such as arthritis and type 2 diabetes. These make it difficult to exercise and my mobility is poor.* [Female participant, 76 years]

XSL•FO
RenderX

In such cases, the issue is not a lack of awareness or willingness, but the incompatibility of technological prompts with physical capacity.

Finally, participants reported that self-efficacy shaped their willingness to use a Fitbit. One male participant expressed:

> *My confidence is very low due to my weight problems. This not only stops me exercising in public but it impacts other areas of my life like learning new skills.* [Male participant, 58 years]

This suggests that psychological and social barriers, such as low confidence and fear of judgment, can limit both physical activity and engagement with new technologies.

Despite these barriers, participants identified several factors that could encourage uptake. Provision of a free device was repeatedly highlighted as a strong motivator:

> *Normally I would not have bought a Fitbit but having had the device provided free has helped motivate me to be more active.* [Female participant, 56 years]

Device provision at registration was seen as a way to reduce cost-related inequalities and promote engagement from the outset.

Participants also emphasized the importance of education and support. One explained:

> *I wasn't sure how best to use the Fitbit. I think a class dedicated on how best to use the Fitbit would have been very useful.* [Female participant, 54 years]

Similarly, health care professionals highlighted the value of hands-on assistance:

> *A number of your study participants needed support to set up the Fitbit device… I could help participants set up their trackers if required.* [Male professional, 34 years]

These accounts underscore the importance of embedding structured technical support into interventions.

Goal setting was another key enabler. A professional suggested:

> *As part of the Weigh to Go programme we could speak to each participant and set them physical activity goals over the 15 weeks.* [Female professional, 41 years]

This reflects established evidence that personalized and incremental goal-setting strategies can enhance motivation and sustain behavior change.

Promotion

Health care professionals and participants proposed several mechanisms for promoting Fitbit use within clinical settings. The most direct approach was integration into consultations:

> *This could be promoted through our practice or during a consultation* [Female professional, 41 years]

Others suggested passive promotion through printed materials displayed in clinical environments:

> *The NHS could produce posters, leaflets and handout documents within their premises or other public buildings.* [Male professional, 34 years]

Finally, national-level campaigns were seen as potentially effective for raising awareness, with 1 participant proposing:

> *A national campaign in local newspapers informing patients of this service might work.* [Male participant, 41 years]

Within the Weigh to Go program, promotion at the point of registration was viewed as crucial:

> *I was informed of this study when I first registered for the Weigh to Go programme. A friend of mine attended a different venue but was not told of this. I think promoting this at each venue would be a good idea.* [Female participant, 69 years]

Beyond clinical settings, community spaces were highlighted as useful sites for engagement:

> *I attend a companion group at my local community center each week… This would be a good place to promote the Fitbit.* [Female participant, 66 years]

Such suggestions reflect the importance of both formal and informal networks in raising awareness.

Clinical Care

Participants emphasized the potential for Fitbit devices to contribute to clinical care. A health care professional highlighted:

> *Using an activity tracker could form part of a patient's care and if properly managed could improve their physical activity levels and improve their health.* [Male professional, 34 years]

This was reinforced by participants who valued the potential for professional monitoring:

> *The Fitbit could be used by doctor or nurse to monitor my activity and give me advice and exercise support if required.* [Female participant, 30 years]

Health care professionals also considered the possibility of prescribing Fitbit devices, akin to medication or medical equipment:

> *Activity trackers could be prescribed by say my doctor…* [Female participant, 56 years]

Remote monitoring was seen as particularly advantageous:

> *These pieces of technology would allow for the remote monitoring of a patient through their activity data and subsequently provide appropriate advice and support.* [Female professional, 51 years]

Referral pathways were also considered, with 1 professional suggesting:

> *If a patient is not undertaking enough exercise through analysis of their Fitbit information we could refer them to a local gym.* [Male professional, 34 years]

XSL•FO
RenderX

Finance

Participants highlighted potential costs for both patients and providers. For patients, the primary expense was the purchase of the Fitbit itself:

> *For me the main cost would be purchasing the Fitbit device.* [Female participant, 48 years]

Others noted that the use of Fitbits necessitated additional costs such as smartphones and internet access:

> *To use the Fitbit I would need to pay for internet access and to download the Fitbit app have a mobile phone.* [Female participant, 60 years]

For health care providers, resource implications extended beyond devices to include staff training and time:

> *The main costs for the health care provider beyond the price of the Fitbit would be staff training and staff time.* [Female professional, 47 years]

Conversely, participants suggested that investment in activity trackers could generate long-term savings through improved health outcomes. One professional explained:

> *If patients increase their physical activity and reduce their sedentary behavior the knock-on effect will be improved health and a reduction in the risk of developing future illnesses.* [Male professional, 34 years]

Another emphasized reduced reliance on costly treatments:

> *The healthier patients are then they will need less medications and expensive treatments.* [Female professional, 51 years]

Fitbit Functions

Step counting was the most commonly used and valued feature, as 1 participant explained:

> *I tended to only use the step function on the Fitbit as I found this easiest to understand.* [Female participant, 56 years]

Other participants valued the ability to track weight:

> *My Weigh to Go instructor told me how to type my weight into the Fitbit. This allowed me to monitor it while I was on the course.* [Female participant, 30 years]

However, several participants described difficulties interpreting the full range of data provided by the device. One commented:

> *I found the Fitbit motivational but other than steps I did not understand much of the information on the app.* [Female participant, 76 years]

This was seen as a missed opportunity, with others suggesting that staff-led educational sessions could bridge this gap:

> *I would have found it useful if our Weigh to Go class instructor had delivered a session explaining how best to use the Fitbit and how to interpret the information recorded.* [Male participant, 41 years]

Effectiveness (Qualitative)

Participants generally perceived Fitbit devices as effective in increasing awareness of their physical activity levels and progress. For example:

> *I enjoyed using the Fitbit. It helped me see any improvements in my activity levels* [Female participant, 48 years]

The device also acted as a motivational prompt:

> *I found the Fitbit motivated me to be more active and encourage me to do something on days when I felt tired.* [Female participant, 60 years]

Health care professionals suggested that activity trackers could be integrated into structured goal-setting strategies within interventions such as Weigh to Go. As 1 professional argued:

> *These devices could certainly be incorporated into the Weigh to Go programme and be used as an effective goal setting tool.* [Male professional, 34 years]

Summary

Overall, the findings highlight both the opportunities and challenges associated with incorporating Fitbit devices into weight management and clinical care. While barriers such as cost, digital literacy, and health conditions were evident, enablers such as device provision, education, and personalized goal-setting enhanced engagement. Promotion strategies at both clinical and community levels were proposed, and participants recognized the potential for activity trackers to support both behavior change and clinical monitoring. However, realizing these benefits will require attention to structural inequalities and adequate resourcing to ensure equitable access and effective integration.

Discussion

Principal Findings

This discussion reports a pragmatic evaluation of integrating a consumer activity tracker (Fitbit Charge 5) into Weigh to Go, a community weight management program delivered across Lanarkshire, Scotland. The purpose was not to test the effectiveness of Fitbit devices per se, but to examine feasibility, acceptability, and implementation issues in a real-world service using the RE-AIM framework [29,39]. Quantitative (Fitbit steps and MVPA minutes) and qualitative (semistructured interviews with Weigh to Go participants and health care professionals) data were integrated using an abductive approach to illuminate why the observed patterns occurred and what they imply for future scale-up [24-27,37].

The quantitative findings demonstrated an increase in daily steps and MVPA during the structured active phase of the program, followed by a decline during the maintenance phase. These patterns reflect the degree to which the tracker was embedded within the program structure: integration was strongest when supported by weekly sessions, staff engagement, and routine accountability, and weaker when participants were expected to self- manage.

XSL•FO
RenderX

The qualitative findings provided insight into how participants and staff experienced the tracker. Participants valued the immediate feedback, goal-setting features, and sense of accountability provided by the device. Staff recognized the potential of trackers to enhance engagement but reported variability in confidence, training, and time available to support device use.

Reach

Consistent with RE-AIM, reach concerns the absolute number, proportion, and representativeness of individuals who participate [29]. Seventy-one adults consented between July 2023 and January 2024; 59 were enrolled, and 33 contributed analyzable device data, with participants predominantly female and White and with a mean age in the mid-50s. Although 705 participants attended the targeted Weigh to Go venues during the recruitment window, the program did not record how many were approached, constraining precise calculation of recruitment rate and representativeness. Based on venue attendance, a conservative estimate suggests that approximately 10% of potentially eligible attendees consented (ie, 71/705), but this cannot be taken as a true denominator because approach rates were not logged. Future evaluations should prospectively track the recruitment path and report reasons for nonparticipation and attrition, in line with RE-AIM and CONSORT (Consolidated Standards of Reporting Trials) style flow logic adapted for pragmatic studies [39].

Qualitatively, participants and health care professionals emphasized that providing devices at no cost enhanced reach, particularly among people facing financial constraints and among men and ethnic minority groups, who are often underrepresented in weight management services [16,17]. However, digital access and digital literacy remained barriers for some older adults and those with limited technology confidence, echoing persistent “digital divide” concerns in community interventions [42].

Effectiveness

Within RE-AIM, effectiveness refers to impact on key outcomes, including potential harms. The within-subject analyses indicated a significant increase in daily steps between week 1 and week 7 (Bonferroni-adjusted, P=.008), with an average change of >5345 steps (reporting group means), whereas no statistically significant differences were observed for the other pairwise comparisons of steps or for weekly MVPA minutes over the same period. These results are presented cautiously as the study design does not permit causal attribution to the tracker or to any single component of Weigh to Go. Step count increases could reflect early engagement with the program as a whole, seasonal/contextual influences, or novelty effects [43], not necessarily the addition of a Fitbit device.

There was a difference between the recorded change in steps and MVPA. However, step counts often increase through additional light-intensity walking and incidental movement (eg, breaks in sedentary time), whereas MVPA requires higher-intensity bouts of physical activity [5,44]. In our qualitative data, participants described finding steps easy to understand and act upon, while feeling unsure how to interpret intensity metrics, suggesting an opportunity for education sessions focused on translating device feedback (active minutes and heart rate zones) into actionable weekly plans. Importantly, the Fitbit served simultaneously as an intervention component and measurement tool, a limitation we acknowledge below, given concerns about construct overlap and potential reactivity [10].

To improve interpretability in future phases, it is recommended that a single family of follow-up contrasts is used, either (1) stepwise (weeks 1-7, 7-15, and 15-20) or (2) baseline-referenced (week 1 vs weeks 7, 15, and 20) with appropriate multiplicity control (eg, Bonferroni or Holm). In this evaluation, the baseline-referenced approach was chosen to avoid mixing strategies, and weeks 1, 7, 15, and 20 were selected a priori to represent early engagement (week 1), midactive phase (week 7), end of active phase (week 15), and early maintenance (week 20), aligning with the program’s delivery structure.

Rather than evaluating program efficacy, our findings illustrate how the tracker functioned as an integrated component of the intervention. The observed increases in activity reflect how participants engaged with the device within the program context, while the later decline highlights the limits of integration when structured support is reduced.

Adoption

Adoption addresses the proportion and representativeness of settings and staff willing to initiate the intervention [29]. Across the service, 22 instructors and additional managers/administrators are engaged in Weigh to Go delivery, with 12 community venues used during the study period (community centers and fitness centers). Staff interviews revealed high conceptual openness to tracker integration but practical uncertainties regarding what exactly to do with device data, how to interpret dashboards meaningfully, and how to troubleshoot common issues (eg, syncing and account management). Although Fitbit devices are consumer-friendly, clinical/community adoption requires confidence in data interpretation, behavior change coaching with device metrics, and pathways for acting on data (eg, tailoring targets and referrals to supervised activity). This aligns with prior literature that wearables can support self-monitoring and feedback, but do not drive behavior change in isolation; professional guidance and integration into a behavioral program are essential [12,45,46].

At the organizational level, adoption is more likely where there is a protocol specifying eligibility, onboarding, data governance, coaching scripts, troubleshooting trees, and criteria for progression and referral (eg, to gym-based supervision). The interviewees explicitly requested such a protocol, which would also facilitate staff training and quality assurance across venues.

Staff adoption varied across roles and settings. Some instructors embraced the tracker as a motivational tool, while others viewed it as an optional add-on. Organizational adoption was influenced by staff training, perceived relevance, and available time. These findings highlight the importance of organizational readiness and clear role expectations when integrating digital tools into community programs.

XSL•FO
RenderX

Implementation

Implementation refers to fidelity, adaptations, costs, and practicalities [29]. The program delivered a 15-week active phase (education and exercise) followed by an optional 15-week maintenance phase. In practice, attendance during the active phase averaged 11.35 (SD 3.63) sessions, while maintenance attendance averaged 1.91 (SD 2.57) sessions. Only 1 participant attended all 30 sessions across active and maintenance phases; 5 submitted Fitbit data beyond 20 weeks. Interviews pointed to competing life demands, waning novelty, device/app literacy challenges, and variable goal-setting support as proximate reasons for drift. These reasons mirror broader evidence that self-monitoring adherence decays without ongoing reinforcement, social support, and personally meaningful goal setting [46-48].

From an economic standpoint, Weigh to Go's reported average delivery cost was US $90.22 per participant for the 15-week active phase and US $180.44 for 30 weeks (active and maintenance), exclusive of devices. The retail cost of a Fitbit Charge 5 was US $136.30 during the study. A simple per-participant estimate for an activity tracker–enabled 15-week program is therefore approximately US $226.60 (program delivery plus device), rising with staff training and technical support. Venue costs were US $2707.65 per venue annually. Scaling to 125 venues at current venue costs implies approximately US $338,456.16 per year for venues alone, plus session delivery (eg, US $1805.10 per 30-week cycle per venue), devices, training, and support infrastructure. Given low maintenance attendance and data incompleteness, such scaling is unlikely to be cost-effective unless adherence can be substantially improved and unit costs reduced (eg, negotiated device pricing, shared equipment pools, or bring-your-own device models with equity safeguards). Prior work suggests that digital components can be cost-effective at scale only when engagement is sustained [49,50].

Two further implementation issues require explicit acknowledgment. First, measurement/intervention confounding arises when the Fitbit is used simultaneously as both the behavioral intervention component (ie, providing feedback, reminders, or activity prompts) and the outcome assessor (ie, recording steps, activity minutes, or heart rate). This dual role introduces a risk of reactivity, whereby participants may alter their behavior simply because they know their activity is being monitored by the device rather than due to the intended intervention mechanisms. In addition, relying on a single tool for both intervention delivery and measurement raises common method bias concerns, as the observed changes in outcomes may partly reflect the measurement method itself rather than genuine behavioral change [51]. Consequently, improvements recorded in physical activity metrics could be artificially inflated, undermining the ability to disentangle whether changes are attributable to the intervention's efficacy, participants' awareness of being monitored, or inherent limitations in the measurement device. Second, data loss occurred. Missing device data arose primarily from account and syncing errors (Google/Fitbit credential issues and intermittent app–phone connectivity), compounded by irregular wear. These problems were exacerbated by external vendor-driven changes, such as Google/Fitbit account migration, which caused unexpected authentication barriers. Syncing failures meant that even when devices were worn, data often did not upload consistently, while irregular wear patterns, charging lapses, and user disengagement further inflated missingness. Such losses created a disproportionate burden on staff, who had to provide troubleshooting, technical support, and device replacements, increasing hidden resource demands. Beyond logistical challenges, these gaps undermined analytic validity; data were systematically missing from lower-engagement participants, limiting statistical power and biasing outcome estimates. Together, these issues highlight that consumer devices, while seemingly low cost, introduce significant risks to data completeness, program fidelity, and equity unless robust technical and contractual safeguards are in place.

Implementation fidelity varied across sites. Some staff consistently introduced and supported tracker use, while others provided minimal guidance. Barriers included limited training, competing demands during onboarding, and variability in participants' digital literacy. These findings underscore the need for structured implementation protocols, staff training, and dedicated time for device support.

Maintenance

Maintenance concerns sustained individual behavior change and the institutionalization of the program [29]. At the individual level, the sharp fall-off after the active phase indicates that behavioral maintenance was weak. Interviews suggest that goal progression, habit formation, and social accountability mechanisms were insufficiently embedded. Evidence indicates that maintenance benefits from graduated goals, implementation intentions, prompted self-regulation, and structured social support [8,19,48]. At the organizational level, maintenance would require a codified protocol, staff capability for data-guided coaching, and routine use of trackers in referral pathways (eg, condition-appropriate activity prescriptions, remote monitoring, or peer groups), as suggested by both participants and health care professionals and supported by remote monitoring literature [13].

Maintenance was limited at both the individual and organizational levels. Participants reported reduced motivation once the structured phase ended, and staff highlighted the need for clearer protocols, long-term funding, and integration with existing monitoring systems. These findings suggest that sustained integration requires ongoing support, resource allocation, and alignment with organizational processes.

Consideration of Participant With Learning Difficulties

During the study, 1 Weigh to Go participant was identified as having learning difficulties through the demographics survey administered via Qualtrics. Ethical and methodological considerations were taken into account to ensure that their inclusion in the research was appropriate, respectful, and did not compromise the quality of the data collected.

For this participant, support was offered, and a member of the research team was available to provide clarification and to check comprehension at each stage.

XSL•FO
RenderX

Data collection procedures were also adjusted to accommodate the participants' needs. Instructions for the use of the activity tracker were delivered slowly, repeated where necessary, and supplemented with practical demonstrations from venue staff. Written guidance was supported by verbal explanations. Regular check-ins were conducted to ensure that the device was being worn correctly and that data were being recorded as intended.

Ethically, it is important to reflect on the implications of including a participant with learning difficulties. Their participation demonstrates the inclusivity of the study design and highlights the feasibility of using wearable technologies across a broader range of populations, including those with additional support needs. It also underlines the importance of tailoring recruitment and data collection procedures to individual capabilities to maximize both participant experience and data quality.

Future research should more systematically address accessibility and inclusivity, ensuring that interventions involving wearable technologies are designed with flexibility to accommodate individuals with varying levels of literacy, cognitive capacity, and health literacy. By doing so, activity tracker interventions can be made more equitable and representative of real-world populations.

In summary, the study accounted for the needs of a participant with learning difficulties by providing adapted materials, additional support during consent and data collection, and regular check-ins to ensure compliance. Reflection on this inclusion highlights the importance of designing interventions and research methodologies that are accessible to diverse populations, which represents an important area for future development in this field.

The Fitbit as Both Measurement Tool and Intervention Component

A key methodological consideration is that the Fitbit acted simultaneously as both a measurement tool and an intervention component. Its feedback, goal-setting, and reminder features inherently influence behavior, meaning that measurement and intervention effects cannot be fully disentangled. Recognizing this dual role is essential for interpreting outcomes and designing future studies that aim to isolate behavioral effects from device-driven reactivity.

Strengths and Limitations

A key strength of this study is its pragmatic design, which examined the integration of a consumer activity tracker into an existing community-based weight management program (Weigh to Go) under real-world service conditions. By embedding the evaluation within a routine community intervention, findings reflect implementation realities, enhancing ecological validity and the relevance of recommendations for service delivery [39]. The use of the RE-AIM framework further strengthened the study by structuring the analysis across RE-AIM domains, facilitating a comprehensive evaluation of both individual and system-level outcomes [29].

Another strength lies in the mixed methods approach, which combined quantitative Fitbit-derived data with qualitative interviews from both Weigh to Go participants and health care professionals. This methodological triangulation provided richer insight into not only whether physical activity increased but also how and why adoption and engagement occurred [24]. The abductive analytic strategy enabled a nuanced interpretation that linked empirical findings with theoretical concepts, capturing unanticipated patterns such as the divergence between step count improvements and minimal changes in moderate-to-vigorous–intensity physical activity [26,31].

The study also benefited from the inclusion of multiple stakeholders, with perspectives gathered from Weigh to Go participants and health care professionals involved in program delivery. This broadened the analysis beyond individual behavior change to include service delivery challenges, data governance considerations, and resource implications, which are critical elements for informing scale-up [46]. Furthermore, the provision of free Fitbit devices increased participation among individuals who might otherwise face financial barriers, supporting equity of access and contributing to insights around reach [16].

A major strength of this study is its real- world context, which enhances ecological validity. The mixed methods design allowed for a nuanced understanding of integration across RE-AIM dimensions. However, substantial Fitbit data loss limited the representativeness of quantitative findings. This limitation is best understood not only as a data quality issue but also as an indicator of integration challenges, particularly around account management, technical support, and participant responsiveness.

Despite these strengths, several limitations must be acknowledged. First, the noncontrolled design precludes causal inference. Increases in daily step counts, particularly between week 1 and week 7, cannot be attributed solely to Fitbit use, as changes may also reflect general engagement with the Weigh to Go program, seasonal influences, or novelty effects [43]. Second, the Fitbit device functioned both as an intervention and as the primary measurement tool, raising concerns about measurement reactivity, construct overlap, and reliance on proprietary algorithms [10,51].

Third, data completeness and quality issues emerged, with missing Fitbit records due to syncing errors, account access problems, and irregular device wear. Such missingness may bias estimates and reduce the statistical power to detect changes. Fourth, maintenance engagement was low, with very few participants providing data beyond 20 weeks and only 1 completing the full 30-week program. This point was not explored during the qualitative interviews, particularly with the participant who completed the full 30-week program. Overall, these limitations restrict conclusions regarding long-term sustainability, which is critical for population-level type 2 diabetes prevention [48].

Fifth, the sample lacked sociodemographic diversity. The participant group was predominantly White, female, and middle-aged, reducing the representativeness of findings and limiting generalizability to more diverse populations, including men, younger adults, and ethnic minority groups. Given evidence of disparities in both type 2 diabetes prevalence and

XSL•FO
RenderX

digital health engagement, future studies should prioritize targeted recruitment and culturally tailored strategies [51].

Finally, while the study provided illustrative economic estimates, these relied on assumptions regarding device costs, staff time, and venue use. A full economic evaluation incorporating cost-effectiveness modeling and sensitivity analyses is needed to assess scalability and sustainability [49,50].

In summary, this study provides valuable real-world evidence on the integration of consumer activity trackers into a community weight management program, highlighting both opportunities and challenges. Its strengths lie in its pragmatic, mixed methods design and stakeholder inclusion, while its limitations include design constraints, data loss, limited diversity, and uncertain long-term outcomes. These considerations should inform future controlled evaluations, independent measurement strategies, and implementation planning.

Clarification of Pragmatic Evaluation

We describe this study as a pragmatic evaluation to emphasize that it examined how the tracker functioned within a real-world service context, rather than testing the efficacy of the Weigh to Go program under controlled conditions. This does not imply a change in study design but clarifies the interpretive lens applied to the findings.

Implications and Conclusions

Within a real-world community program, integrating an activity tracker is feasible and perceived as acceptable by many participants and staff. However, sustained engagement is the principal challenge and a determinant of value for money. The quantitative signal of increased steps early in the active phase, without corresponding MVPA change, is consistent with light-intensity activity displacement rather than structured intensity gains. Education on interpreting intensity metrics, goal progression, and coaching appears necessary to shift MVPA. To support adoption and implementation at scale, the service should develop a formal protocol covering onboarding, data governance, coaching scripts, and referral triggers, deliver role-specific staff training on data interpretation, troubleshooting, and behavior-change techniques, and strengthen maintenance via scheduled check-ins, social/peer support, graduated goals, and remote prompts/feedback. Economically, large-scale roll-out (eg, 125 venues) will remain fragile unless adherence improves and unit costs, notably device and support, are reduced through procurement and design efficiencies.

Reframed as a pragmatic evaluation rather than an efficacy test, this study contributes evidence on how and under what conditions activity trackers can be integrated into community weight-management services for adults at risk of or living with type 2 diabetes.

Future research should include comparative designs (eg, stepped-wedge or cluster randomized control trials), independent activity assessment (accelerometry), and economic evaluation, coupled with qualitative process evaluations grounded in reflexive thematic analysis to explain mechanisms and context. Only with stronger maintenance supports and clear operational protocols is tracker-enabled scale-up likely to achieve meaningful population health. A key methodological insight is that the Fitbit acted simultaneously as a measurement tool and an intervention component. Its feedback and goal-setting features inherently influence behavior, meaning that measurement and intervention effects cannot be fully separated. Future research should consider designs that explicitly address this dual role.

Funding

The authors declare no financial support was received for this work. No external financial support or grants were received for this work.

Conflicts of Interest

None declared.

References

1. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008 Nov;88(11):1254-1264. [doi: 10.2522/ptj.20080020] [Medline: 18801858]
2. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
3. How many people in the UK have diabetes? Diabetes UK. 2025. URL: https://www.diabetes.org.uk/about-us/about-the-charity/our-strategy/statistics [accessed 2026-03-10]
4. Kennerly A, Kirk A. Physical activity and sedentary behaviour of adults with type 2 diabetes: a systematic review. Practical Diabetes 2018 May;35(3):86. [doi: 10.1002/pdi.2169]
5. Colberg SR, Sigal RJ, Yardley JE, et al. Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. Diabetes Care 2016 Nov;39(11):2065-2079. [doi: 10.2337/dc16-1728] [Medline: 27926890]
6. Delevatti RS, Bracht CG, Lisboa SDC, et al. The role of aerobic training variables progression on glycemic control of patients with type 2 diabetes: a systematic review with meta-analysis. Sports Med Open 2019 Jun 7;5(1):22. [doi: 10.1186/s40798-019-0194-z] [Medline: 31175522]

XSL•FO
RenderX

7. Sardinha LB, Magalhães JP, Santos DA, Júdice PB. Sedentary patterns, physical activity, and cardiorespiratory fitness in association to glycemic control in type 2 diabetes patients. Front Physiol 2017;8:262. [doi: 10.3389/fphys.2017.00262] [Medline: 28503154]
8. Avery L, Flynn D, van Wersch A, Sniehotta FF, Trenell MI. Changing physical activity behavior in type 2 diabetes: a systematic review and meta-analysis of behavioral interventions. Diabetes Care 2012 Dec;35(12):2681-2689. [doi: 10.2337/dc11-2452] [Medline: 23173137]
9. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57. [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
10. White JW 3rd, Finnegan OL, Tindall N, et al. Comparison of raw accelerometry data from ActiGraph, Apple Watch, Garmin, and Fitbit using a mechanical shaker table. PLoS One 2024;19(3):e0286898. [doi: 10.1371/journal.pone.0286898] [Medline: 38551940]
11. Bender MS, Cooper BA, Park LG, Padash S, Arai S. A feasible and efficacious mobile-phone based lifestyle intervention for Filipino Americans with type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017 Dec 12;2(2):e30. [doi: 10.2196/diabetes.8156] [Medline: 30291068]
12. Lim S, Kang SM, Kim KM, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2016 Apr;53(2):189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]
13. Shei RJ, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? Eur J Appl Physiol 2022 Sep;122(9):1975-1990. [doi: 10.1007/s00421-022-04951-1] [Medline: 35445837]
14. Weigh to Go. NHS Lanarkshire. 2024. URL: https://www.nhslanarkshire.scot.nhs.uk/services/weight -management-service/weigh-to-go/ [accessed 2026-03-10]
15. Dixon D, Johnston M. MAP: A mnemonic for mapping BCTs to three routes to behaviour change. Br J Health Psychol 2020 Nov;25(4):1086-1101. [doi: 10.1111/bjhp.12458] [Medline: 32715603]
16. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
17. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
18. Pourzanjani A, Quisel T, Foschini L. Adherent use of digital health trackers is associated with weight loss. PLoS ONE 2016;11(4):e0152504. [doi: 10.1371/journal.pone.0152504] [Medline: 27049859]
19. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289. [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
20. Ellingson LD, Lansing JE, DeShaw KJ, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019 Feb 14;7(2):e10988. [doi: 10.2196/10988] [Medline: 30762582]
21. Hodkinson A, Kontopantelis E, Adeniji C, et al. Interventions using wearable physical activity trackers among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2021 Jul 1;4(7):e2116382. [doi: 10.1001/jamanetworkopen.2021.16382] [Medline: 34283229]
22. Dehghan Ghahfarokhi A, Vosadi E, Barzegar H, Saatchian V. The effect of wearable and smartphone applications on physical activity, quality of life, and cardiovascular health outcomes in overweight/obese adults: a systematic review and meta-analysis of randomized controlled trials. Biol Res Nurs 2022 Oct;24(4):503-518. [doi: 10.1177/10998004221099556] [Medline: 35535558]
23. Lincoln YS, Guba TM. Naturalistic Inquiry: Sage Publications; 1985. URL: https://uk.sagepub.com/en-gb/eur/naturalistic-inquiry/book842 [accessed 2026-03-17]
24. Patton MQ. Qualitative Research & Evaluation Methods, 4th edition: Sage Publications; 2015. URL: https://uk.sagepub.com/en-gb/eur/qualitative-research-evaluation-methods/book232962 [accessed 2026-03-17]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
27. Braun V, Clarke V. Thematic Analysis: A Practical Guide: Sage Publications; 2021. URL: https://us.sagepub.com/en-us/nam/thematic-analysis/book248481 [accessed 2026-03-17]
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
29. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]

XSL•FO
RenderX

30. Shorten A, Smith J. Mixed methods research: expanding the evidence base. Evid Based Nurs 2017 Jul;20(3):74-75. [doi: 10.1136/eb-2017-102699] [Medline: 28615184]
31. Tavory I, Timmermans S. Abductive Analysis: Theorizing Qualitative Research: The University of Chicago Press; 2012. URL: https://press.uchicago.edu/ucp/books/book/chicago/A/bo18785947.html [accessed 2026-03-17]
32. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
33. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Jul;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8]
34. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
35. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
36. Mark M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual Soc Res 2010;11(3) [FREE Full text]
37. Timmermans S, Tavory I. Theory construction in qualitative research: from grounded theory to abductive analysis. Sociol Theory 2012;30(3):167-186. [doi: 10.1177/0735275112457914]
38. Peirce CS. The Essential Peirce: Selected Philosophical Writings: Indiana University Press; 1998. URL: https://www.jstor.org/stable/j.ctt16gz4vr [accessed 2026-03-17]
39. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
40. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1) [FREE Full text]
41. Tracy SJ. Qualitative quality: eight “Big-Tent” criteria for excellent qualitative research. Qual Inq 2010 Dec;16(10):837-851. [doi: 10.1177/1077800410383121]
42. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40. [doi: 10.2196/mhealth.4461] [Medline: 27122452]
43. Ledger D, McCaffrey D. Inside wearables: how the science of human behavior change offers the secret to long-term engagement. : Endeavour Partners; 2014.
44. Tudor-Locke C, Craig CL, Beets MW, et al. How many steps/day are enough? for children and adolescents. Int J Behav Nutr Phys Act 2011 Jul 28;8(1):78-78. [doi: 10.1186/1479-5868-8-78] [Medline: 21798014]
45. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
46. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
47. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164. [doi: 10.2196/mhealth.7311] [Medline: 29084709]
48. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
49. Wang JB, Cataldo JK, Ayala GX, et al. Mobile and wearable device features that matter in promoting physical activity. J Mob Technol Med 2015;5(2):2-11. [doi: 10.7309/jmtm.5.2.2]
50. Dombrowski SU, Knittle K, Avenell A, Araújo-Soares V, Sniehotta FF. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. BMJ 2014 May 14;348:g2646. [doi: 10.1136/bmj.g2646] [Medline: 25134100]
51. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
COREQ: Consolidated Criteria for Reporting Qualitative Research
MVPA: moderate-to-vigorous–intensity physical activity
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

XSL•FO
RenderX

Edited by R Zhao; submitted 08.Jan.2026; peer-reviewed by K Winfree, P Kelly; revised version received 18.Feb.2026; accepted 19.Feb.2026; published 25.Mar.2026.

Please cite as:
Hodgson W, Kirk A, Lennon M, Janssen X, Kennedy D
Exploring the Integration of Consumer Activity Trackers Into a Community Weight Management Intervention to Support Physical Activity in Adults at Risk for or With Type 2 Diabetes: Mixed Methods Study Using the RE-AIM Framework
JMIR Diabetes 2026;11:e91073
URL: https://diabetes.jmir.org/2026/1/e91073
doi:10.2196/91073

© William Hodgson, Alison Kirk, Marilyn Lennon, Xanne Janssen, David Kennedy. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 25.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cultural and Technological Barriers to Telehealth Adoption for Type 2 Diabetes Management Among Asian American Patients: Qualitative Case Study

Devi Gurung States[1,2], MS, MSW, MPH, DHSc, PhD

[1]Himalayan Family Healthcare Project (HFHCP), 4145 Manchester Ave, St. Louis, MO, United States

[2]Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Devi Gurung States, MS, MSW, MPH, DHSc, PhD
Himalayan Family Healthcare Project (HFHCP), 4145 Manchester Ave, St. Louis, MO, United States

Abstract

Background: In the past decade, telehealth has transformed health care delivery by allowing patients more rapid and convenient access to necessary care without the cost and logistical challenges of traveling to a health care facility. Telehealth services can benefit patients with type 2 diabetes mellitus (T2DM) amid a growing epidemic of T2DM in the United States that affects people of all ages and races. In 2020, 33 million people were diagnosed with this chronic disease, with the number expected to rise by 50% by 2040. Telehealth facilitates regular contact between patients and their providers, especially when there are geographic barriers and time constraints prohibiting physical interaction, at little or no added cost to the patient and at their convenience.

Objective: This study examines cultural and technological barriers affecting telehealth adoption among Asian American people with T2DM.

Methods: A qualitative case study approach was employed, utilizing semistructured interviews with 30 Asian American individuals in Missouri. Thematic analysis was used to identify key barriers.

Results: Four major barriers emerged: (1) language and cultural barriers—limited availability of translated materials and interpreters; (2) limited digital literacy and access—older adults and individuals with low technological exposure struggled with telehealth platforms; (3) limited provider recommendations—health care providers did not actively endorse telehealth, reducing patient awareness of telehealth as an option; and (4) technology access and infrastructure disparities—low-income participants faced challenges with the costs of and access to broadband and telehealth-compatible devices.

Conclusions: Addressing cultural and technological barriers is crucial to increasing telehealth adoption among Asian American people with T2DM. Culturally tailored interventions, provider engagement, and digital literacy programs should be prioritized. Policy efforts must focus on expanding broadband access and providing multilingual telehealth resources.

(JMIR Diabetes 2026;11:e75689) doi:10.2196/75689

KEYWORDS

Asian American; type 2 diabetes; telehealth; digital literacy; cultural barriers; health disparities

Introduction

Telehealth refers to a tool, process, or system that can provide patients with a simpler appointment scheduling process, remote access to clinical services that lessen the need for travel, and increased interaction with health care providers to improve medical outcomes while reducing health care costs [1,2]. In the past decade, telehealth has transformed health care delivery by allowing patients more rapid and convenient access to necessary care without the cost and logistical challenges of traveling to a health care facility [3]. Telehealth services can benefit patients with type 2 diabetes mellitus (T2DM) amidst a growing epidemic of T2DM in the United States that affects people of all ages and races [4]. In 2020, 33 million people were diagnosed with this chronic disease, with the number expected to rise by 50% by 2040 [4].

Telehealth facilitates regular contact between patients and their providers, especially when geographic barriers and time constraints prohibit physical interaction, at little or no added cost to the patient and at their convenience [5]. Telehealth services have the potential to optimize T2DM management by empowering patients to engage in self-care to slow disease progression, prevent complications, and lessen the health care burden [4,6]. Self-management among patients with T2DM includes proper nutrition, adequate physical activity, regular blood glucose monitoring, medication compliance, disease knowledge, lifestyle modifications, and self-efficacy [7]. To ensure effective self-management among patients at high risk

XSL•FO
RenderX

or with a diagnosis of T2DM, telehealth can facilitate digital health coaching for long-term management or prevention of T2DM across population subgroups [8]. Asian American people include multiple subgroups (eg, Chinese, Indian, Vietnamese, Filipino, Korean, and Nepali). Cultural differences among these groups are substantial, especially among first-generation individuals.

According to the US Department of Health and Human Services [9], Asian American people are 40% more likely to be diagnosed with diabetes than non-Hispanic White individuals. Despite this higher prevalence, telehealth utilization among Asian American people remains low, as this population is disproportionately underserved [6,10]. Several factors contribute to this disparity, including cultural norms, digital literacy, provider engagement, and technological access [11]. Many Asian American communities face language barriers, limiting their ability to navigate telehealth platforms effectively [12,13], and older adults may struggle with technological proficiency, creating challenges in virtual health care engagement.

Given the disproportionate burden of T2DM among Asian American people and their persistently low telehealth usage, this study sought to answer the following questions: what cultural and technological barriers limit adoption? and what strategies might address them? These questions are urgent given the stakes, including T2DM complications, loss of care access, and widening racial health disparities. By centering user experience and structural constraints, the study identifies critical leverage points for more equitable digital health policy and practice.

Understanding these barriers is critical for designing equitable telehealth interventions that improve access to diabetes care. Therefore, this study explored the cultural and technological challenges hindering telehealth adoption among Asian American people with T2DM, providing insights for health care providers and policymakers. The guiding theoretical framework was the unified theory of acceptance and use of technology (UTAUT), an integrated approach used to predict and understand how and why individuals or groups accept or reject various technologies [11]. In the context of this inductive study, UTAUT provided a foundation for examining telehealth adoption among Asian American people with T2DM and was used to scaffold the study. UTAUT was used to identify several characteristics that influence the adoption of health technologies, which included social influence, performance expectancy, effort expectancy, facilitating conditions, privacy risk, and the threat of the disease from which the patient is ailing. By investigating factors affecting telehealth use within this population, specifically for the management of T2DM, a greater understanding of barriers to adoption of telehealth services and associated effects was achieved. Depending on the applicability of this study's findings to specific Asian American communities, this knowledge can be applied to support the development of more effective methods of tailoring telehealth services to meet the unique needs of Asian American people, as well as those of other underserved populations. This study provided insights regarding ways to incorporate culturally relevant telehealth approaches to T2DM management into mainstream health care practice.

In addition to the UTAUT framework, this study was informed by a critical health equity lens that emphasizes how structural inequities—such as limited English proficiency, systemic underinvestment in minority-serving institutions, and digital exclusion—shape health care access. This critical perspective extends UTAUT by interrogating not only users' perceptions of telehealth but also the sociotechnical structures that enable or constrain its use among marginalized populations. By situating Asian American T2DM management within broader systems of racialized health care access, this study provides an intersectional understanding of telehealth adoption.

Methods

Study Design and Researcher Perspective

A qualitative case study approach was used to explore telehealth adoption barriers among Asian American people with T2DM in Missouri. The purpose of the study was to examine Asian American people's perspectives to determine the factors that influence their adoption of telehealth for T2DM management. Regarding positionality, my interest in this demographic is that I am an Asian American. Subjective perspectives, rather than objective facts, were elicited from participants; therefore, a qualitative methodology was appropriate, as it enabled in-depth examination [14].

The study focused on Asian American people with T2DM in Missouri due to the group's elevated diabetes risk and underrepresentation in telehealth research. Missouri, as a Midwestern state with rising Asian American populations [15], offered a novel geographic and demographic context. A qualitative case study design was used to capture the complexity of individual, cultural, and technological factors shaping telehealth use, particularly within a population that often faces intersecting language and access barriers. This approach aligns with case study methodology's strengths in revealing nuanced insights within specific, bounded systems.

The study population included Asian American people with T2DM via convenience sampling [16]. Convenience sampling may overrepresent individuals with similar socioeconomic status or community ties and introduce selection bias. Inclusion criteria specified that participants be Asian American, 21 years or older, and diagnosed with T2DM. Those with type 1 diabetes mellitus and gestational diabetes were excluded, along with anyone whose health condition was not under the management of a health care provider. Previous experience with telehealth was not necessary.

Participants were recruited through community health organizations. Patients receiving care from home care agencies in Missouri, being under the direct supervision of health care providers, were asked to volunteer for the study. Participants were categorized into 2 groups: adopters (those who had adopted telehealth for self-management) and nonadopters (those who had not adopted telehealth for self-management).

Ethical Considerations

This study was reviewed and approved by the Capitol Technology University Institutional Review Board (approval IRB05242023a, approved on June 8, 2023). All participants

XSL•FO
RenderX

provided informed consent prior to participation. Privacy and confidentiality were protected by institutional review board approval; data were deidentified and securely stored. No compensation was provided.

Data Collection

A pilot study was conducted to establish the reliability of the semistructured interview, with 10 participants selected from the study population. These participants were then excluded from the main study. Data collected during the pilot study were scrutinized to determine the instrument's capacity to collect data relevant and applicable to the study aims. The findings of the pilot study were used to review the data collection instrument and processes and implement any necessary instrument modifications to enhance its reliability.

Data were collected via semistructured interviews [17] with an open-ended question format. The following are sample questions:

1. (AU) Have you used telehealth services for diabetes management before? If yes, could you describe your experience with using telehealth? If not, can you explain the reasons for your decision not to use it?
2. (PE) How do you think telehealth can be useful for managing type 2 diabetes? Please explain.
3. (PE) What are the benefits of using telehealth for diabetes management?

Using a flexible interview schedule, participants engaged in these interviews through phone and video calls, which were digitally audio-recorded. Probes, follow-up questions, and comments were used to encourage participants to clarify statements where further information was required for a comprehensive understanding of experiences they described [17]. Through phone conversations and video calls, participants shared their experiences, challenges, and perceptions regarding the use of telehealth, especially those associated with management of T2DM.

Data Demographics

Data gathered during the semistructured interviews included participants' demographic information: race, sex, household size, household income, occupation, education level, and age. Data were collected on performance expectancy, indicating perspectives on their expectations of how telehealth might be helpful for the management of T2DM and the realization of health goals. Furthermore, data on effort expectancy were gathered, describing the participants' perceptions of the ease of use of telehealth and factors they believed would affect the ease of use negatively and positively. Finally, data on the social and cultural influences on the use of telehealth for managing T2DM were collected, examining participants' social networks (eg, family members, friends, and health care providers) and cultural factors (eg, preference for in-person care and language barriers). Data regarding participants' personal innovativeness, trust, behavioral intention, and actual use of telehealth services were also gathered. Personal innovativeness pertains to their comfort levels in using new technologies such as telehealth, while trust refers to their level of confidence in the security and privacy offered by telehealth applications and whether they had reservations about sharing their health information online. Behavioral intentions examined whether the participants would adopt telehealth for the management of T2DM and the factors that may influence those decisions. Finally, participants' actual use of telehealth in the past was explored, with data collected on their associated experiences. For participants who did not have previous experience using telehealth, data were collected on their willingness to use it in the future.

Data Analysis

Thematic analysis was employed to identify recurring patterns in participant responses. Qualitative data were categorized based on patterns that form specific themes; data excerpts were thereby organized within the concepts outlined in the UTAUT theoretical framework [18-20]. We first conducted line-by-line inductive coding of all transcripts to identify emergent concepts. These codes were grouped deductively under UTAUT constructs and refined through axial coding into subthemes. Finally, overlapping subthemes were consolidated into 4 primary barrier categories. Qualitative data from the interviews were examined using axial coding, enabling codes, subcategories, and categories contained in the participants' perspectives to be more easily identified [19,21].

Results

Thematic refinement yielded 4 consolidated barriers to telehealth adoption (language and cultural barriers, digital literacy and access, limited provider recommendations, and technology/infrastructure disparities), derived from 9 initial themes.

Data Demographics

A total of 30 participants were recruited for the study: 15 adopters and 15 nonadopters to telehealth services. The 15 adopters were between the ages of 31 and 57, with the average age being 35 years. All of the adopters had attended institutions of higher education, with 6 having bachelor's degrees, 4 having master's degrees, and 2 having doctorates. All of the adopters were Asian American individuals, with multiple ethnicities represented, including Chinese, Indian, Vietnamese, Filipino, Korean, and Nepali. Men were disproportionately represented, with 12 men compared to only 3 women. The highest household income was US $200,000, while the lowest was US $36,000, with most households consisting of 3 people. Participants who were telehealth adopters also worked in various professions, including information technology, academia, and engineering, as well as in miscellaneous jobs (Table 1).

XSL•FO
RenderX

Table . Demographics of the adopters.

Adopter No.	Household size	Age (y)	Ethnicity	Gender	Household income (US $)	Occupation	Education level
1	4	45	Nepali	Male	200,000	IT professional	Master's
2	4	39	Chinese	Male	90,000	Auto mechanic	Bachelor's
3	5	38	Chinese	Male	150,000	Researcher	Bachelor's
4	3	35	Indian	Male	80,000	IT professional	Bachelor's
5	4	51	Korean	Female	100,000	Computer science	Bachelor's
6	3	57	Korean	Male	150,000	IT networking	College
7	4	44	Vietnamese	Male	160,000	Professor	PhD
8	1	36	Filipino	Male	36,000	Sushi cook	College
9	4	46	Vietnamese	Female	120,000	Hair stylist	College
10	3	44	Chinese	Female	110,000	Academia	PhD
11	3	31	Vietnamese	Male	50,000	IT	Bachelor's
12	2	35	Filipino	Male	75,000	Retired army personnel	Bachelor's
13	4	49	Nepali	Male	175,000	Aircraft engineer	Master's
14	3	45	Indian	Male	150,000	Civil engineer	Master's
15	3	35	Indian	Male	108,000	Data engineer	Master's

The 15 nonadopters of telehealth services were between the ages of 32 and 80, with the average age being 55 years. Participants in this group were less educated in comparison to the adopter group, with only 1 having a graduate degree. Five nonadopters graduated high school, while 9 had no formal education. However, as with the adopter group, there were various Asian American ethnicities represented in the group of nonadopters, including Nepali, Indian, Chinese, Filipino, Vietnamese, and Korean. There was less gender disparity among nonadopters compared to adopters, with 9 females and 6 males. The nonadopters earned much less than the adopters in terms of household income, which was between US $9000 and US $36,000. Most households had 2 people. Finally, 10 of the nonadopters were unemployed, while only 3 worked part-time (see Table 2).

Table . Demographics of nonadopters.

Nonadopter No.	Household size	Age (y)	Ethnicity	Gender	Household income (US $)	Occupation	Education level
1	1	80	Korean	Female	9000	Unemployed	High school
2	2	45	Korean	Female	12,000	Casual labor	High school
3	2	75	Chinese	Female	9000	Casual labor	High school
4	1	67	Indian	Male	9000	Unemployed	High school
5	2	62	Filipino	Female	11,400	Unemployed	No education
6	3	55	Nepali	Female	10,000	Unemployed	No education
7	3	67	Nepali	Male	11,000	Unemployed	No education
8	2	53	Indian	Female	12,000	Unemployed	No education
9	2	61	Vietnamese	Male	10,000	Unemployed	No education
10	2	38	Nepali	Male	15,000	Casual labor	No education
11	2	55	Vietnamese	Male	11,000	Unemployed	No education
12	3	32	Chinese	Female	36,000	Student	Graduate
13	2	36	Filipino	Male	36,000	Cook	High school
14	2	58	Indian	Female	12,000	Unemployed	No education
15	2	52	Nepali	Female	12,000	Unemployed	No education

XSL•FO
RenderX

Though past researchers have found facilitators to incorporating telehealth, this study only yielded barriers. To review, facilitators of telehealth implementation were videoconferencing, caregiver engagement, and delivery via the favored language of patients and caregivers [22]. Approaches to enhance telehealth consultations included in-person meetings to establish a relationship before shifting to telehealth and using text and audio telemonitoring to ensure that patients understood advice and instructions [22].

The interpretation following thematic analysis enabled the identification of 4 barriers to telehealth adoption among Asian American individuals with T2DM: (1) language and cultural barriers; (2) digital literacy and access; (3) limited provider recommendations; and (4) technology and infrastructure disparities. Beyond barriers, participants highlighted facilitators that can inform targeted implementation strategies, including provider recommendations, interpreter support, device access, insurance coverage, and perceived convenience.

Language and Cultural Barriers

Nonadopters reported several problems and challenges with telehealth systems, with the most prevalent problem being a language barrier. Due to the lower level of education among nonadopters and most being first-generation immigrants to the United States, many could not speak English. The language barrier was an obstacle unless they found a doctor with whom they shared a common language. In fact, 10 of the 15 nonadopters spoke of the language barrier as the main challenge affecting their use of telehealth.

Digital Literacy and Access

Effort expectancy involves the convenience and usability levels that adopters of telehealth experience when using the system [23,24]. Participants' perceptions of the ease or difficulty of use of telehealth and any associated problems were explored, and a learning curve associated with initial use of telehealth services was identified.

Several adopters noted that there was a steep learning curve and they initially experienced significant difficulties. However, they quickly added that after using telehealth several times and/or having a doctor or a proficient family member explain its use, it eventually became easier to navigate. Some adopters, such as information technology professionals, did not have any problems using the telehealth system. Nonetheless, adopters mentioned some problems with the use of telehealth, such as not having access to the internet. Some telehealth systems are more complex than a simple call to the doctor and require the use of mobile apps accessible only through a smartphone connected to the internet. This issue affects accessibility, especially if the telehealth appointment is scheduled for a time when internet access is limited or not possible due to travel or other factors. Another problem associated with the use of telehealth relates to the availability of supporting technology, such as smartphones or other devices capable of complex operations (eg, camera-enabled desktop computers or laptops). Lack of these technologies represents a significant barrier to using telehealth.

Limited Provider Recommendations

Though provider recommendations were limited, some participants did receive recommendations. Some adopters were forced by circumstances to use telehealth to aid in managing their diabetes, as was the case for P15:

> *I was able to consult with a doctor over the phone to discuss my diabetes. I think useful; you can actually ask questions about your diabetes issues you have at any time.*

Some of those circumstances included constraints on health care access due to the COVID-19 pandemic, like P5, who said:

> *My doctor also asked me to use telehealth. My doctor always communicates with me via text about my health issues. I think my doctor's advice influenced me quite a bit. Because a doctor is a medical doctor, she knows what she's doing. And she was strongly recommended, especially during COVID-19 time.*

Although they were required to use telehealth, the positive experience of doing so encouraged them to continue after the pandemic.

Technology and Infrastructure Disparities

None of the nonadopters had ever used telehealth to assist them in their management of T2DM. There were various reasons given for nonuse, ranging from a lack of awareness of the existence of telehealth systems to not knowing how to use the system, like P2:

> *I think training about telehealth and how to use it.*

However, some nonadopters also mentioned facing language barriers, which prevented them from using telehealth, like P7:

> *No one said anything because of my language barrier.*

Performance Expectancy

Performance expectancy is the degree to which a user believes that using telehealth will help them make gains in managing their health, for this study, T2DM specifically. Performance expectancy was assessed through examining participants' thoughts on how using telehealth could help with their T2DM, as well as the perceived benefits of using telehealth.

Many adopters thought telehealth would introduce convenience and flexibility into the management of their T2DM because it eliminated travel to the doctor's office. In addition, cost-effectiveness and time saved were mentioned by adopters as expected benefits of using telehealth. For example, adopters felt its use was cost-effective, saving them a trip to the doctor—and the associated expenses—while its capacity to provide secure and rapid access to health services ultimately saved valuable time. Adopters reported that using telehealth enhanced their access to health care services.

For nonadopters, performance expectancy was significantly lower compared to adopters. Specifically, there was a significant number of nonadopters who were not aware that telehealth might enhance diabetes management, perhaps because some were not familiar with telehealth systems. Nonetheless, many believed that using telehealth would be beneficial in helping them to

XSL•FO
RenderX

manage their T2DM more effectively. Furthermore, when asked about the benefits of using telehealth, several nonadopters highlighted the convenience of not having to travel to the doctor's office, reduced cost, and time efficiency as likely benefits. These benefits were similar to those mentioned by adopters. However, some nonadopters did not perceive that there would be any benefits associated with the use of telehealth, and their overall preference for in-person care was considerable. See Table 3 for adopters' views on performance expectancy.

Table . Selected quotes by adopters about performance expectancy.

Participant	Adopter/nonadopter	Selected quotes
1	Adopter	But you still have the benefit of using telehealth. For example, it's flexible because you do not have to travel to your doctor's office, saving time.
3	Adopter	It is beneficial because I can call my doctor from anywhere. It is a benefit because I cannot travel and receive timely care for my problems.
5	Adopter	It was pretty seamlessly easy, I thought. Because I asked for the appointment, they asked me to fill out the information online, which I did pretty quickly. And then, about an hour later, they said they would call me back with the doctor. And so I got connected with the doctor, and I was able to consult with a doctor over the phone to discuss my diabetes. I think useful; you can actually ask questions about your diabetes issues you have at any time. Even though you're not in the town, because sometimes I'm not in the United States. I am in Korea, visiting my parents or my friends in Korea I'm there for maybe three months or four months. I can just call or email my doctor. I can always connect with my doctor online or on the phone about my diabetes, and I don't have any problem with that.
2	Adopter	Like I said, it's convenient because I can call my doctor anywhere from a distance. It helps with remote monitoring and consultation of blood sugar levels. It is cost-saving because the gas price is going too high.
15	Adopter	I am very busy, and I do not have to drive to the doctor's office. So it saves money and time. It's a convenience.
13	Adopter	Like I said, using telehealth is helping me access the doctor faster. For example, it takes more than two months to get an appointment for the office visit, but it takes less than one week to make a telehealth appointment.
8	Adopter	Besides my diabetes problems and other health problems, telehealth increases access to my care in terms of fast service.
11	Adopter	It does help me to monitor my blood sugar. Using telemedicine increases my chances of improving my diabetes. I don't have to always wait for your appointment.
12	Adopter	Honestly, it has been a lot more accessible, especially during the pandemic time. I can get a telehealth appointment within a week or so, but office visits take more than one month for an appointment.

Data Analysis

The following 9 themes were identified using thematic analysis: actual use, performance expectancy, effort expectancy, social influence, facilitating conditions, cultural influence, personal innovativeness, trust, and behavioral intention.

XSL•FO
RenderX

Actual Use

Information was collected on whether participants had used telehealth to talk to their physicians about diabetes, their experiences, and, among those who reported not having used telehealth, the reasons for not using it. Findings showed that there was a high usage of telehealth among adopters, who felt that telehealth was beneficial. For example, P13 said:

> *Like I said, using telehealth is helping me access the doctor faster. For example, it takes more than two months to get an appointment for the office visit, but it takes less than one week to make a telehealth appointment.*

Similarly, P15 said:

> *I am very busy, and I do not have to drive to the doctor's office. So it saves money and time. It's a convenience.*

This group reported positive experiences that centered on the benefits of telehealth, including its convenience. Table 4 shows adopters' views on actual use.

See Table 5 for a summary of nonadopters' views on actual use.

Table . Selected quotes by adopters about actual use.

Participant	Adopter/nonadopter	Selected quotes
1	Adopter	I've used telehealth before for my diabetes management. I feel that telehealth is beneficial. I think it's more convenient, easy to use, and as you start using it more and more.
4	Adopter	I have talked to my doctor by phone about my diabetes care. One thing I like about that is I don't have to wait a month for the doctor. I can call them whenever I need to and discuss it with the doctor. It's convenient that I do not have to wait too long.
11	Adopter	I have talked to my doctor before. Usually, either be on the phone or a Zoom call. It was easy for me after COVID-19; everything was on lock-down. So there is no choice.
10	Adopter	I have been using telehealth since the COVID-19 pandemic. Because I still feel the risk of exposure to the virus

Table . Selected quotes by nonadopters about actual use.

Participant	Adopter/nonadopter	Selected quotes
1	Nonadopter	I don't know much about it. I never talked to my doctor using the phone or anything because I did not know there was a telehealth service available. I have no idea. My doctor did not say anything to me.
6	Nonadopter	I do not use it because of language barrier.
5	Nonadopter	I do not know how to use it. I never knew about it.

Social Influence

Social influence is the degree to which actors in a person's social circles, such as the primary physician, family members, and friends, influence the decision of an individual to use telehealth. Participants were asked whether their doctors, friends, or family members encouraged the use of telehealth; they were then asked to describe the nature of that input.

Among adopters, the social influence of doctors and family members impacted their decision to start using telehealth. As doctors and family members communicated the benefits of using telehealth services, participants were led to consider it. When doctors described the benefits of convenience, cost-effectiveness, and time efficiency, they were influential in the participants' decision. In addition to doctors, adopters indicated that family members also had an important role in encouraging them to adopt telehealth. Family members of participants indicated that there were tremendous benefits of using telehealth for managing diabetes and described it as an effective health care service that is gaining traction for treating this chronic illness.

In contrast, nonadopters did not receive advice from their doctors, friends, or family members regarding the use of telehealth. Some of the nonadopters believed that their nonproficiency in English was part of the reason their doctors never recommended telehealth. Others stated that their physicians did not suggest it.

XSL•FO
RenderX

Facilitating Conditions

Facilitating conditions concerns the user's belief that the technical infrastructure and other conditions necessary for the use of telehealth already exist [25]. Participants' perceptions regarding whether the necessary resources to facilitate telehealth use for diabetes care were in place as well as their perceptions regarding whether they had access to necessary devices were examined.

Regarding owning the necessary devices, all adopters reported that they owned a smartphone, allowing them to easily connect to the internet and run the mobile apps necessary to use telehealth effectively. In addition, insurance coverage was mentioned by adopters as a necessary support enabling the use of telehealth. Several adopters reported that their insurance did not cover the service and that they had to pay out-of-pocket. This finding highlighted the need for insurance coverage as one of the facilitating conditions to enable adopters to use telehealth services when needed for managing diabetes. Adopters also reported that they would like support in the form of training on effective use of telehealth in general, as well as for diabetes management, specifically. There was a perception that doctors and their staff could also benefit from such training.

Nonadopters faced conditions that did not facilitate their use of telehealth. For instance, a majority of nonadopters did not own the devices, such as smartphones or laptops, required for telehealth access. In addition, other nonadopters deemed the services of a translator to be an important support resource necessary for them to use telehealth services effectively. This finding was logical, considering that 10 of the 15 nonadopters identified language as a barrier limiting their use of telehealth. Furthermore, most nonadopters pointed out that financial health was an important aspect for them to consider in relation to telehealth services. Some reported social security as their sole source of income. Those who indicated that financial assistance was necessary to support their use of telehealth services reported they would use such aid to purchase a phone and pay for the services of an interpreter. In addition to financial aid, they reported a need for training on telehealth and how it applies to diabetes management. See Table 6 for a summary of nonadopters' views on facilitating conditions.

Table . Selected quotes by nonadopters about facilitating conditions.

Participant	Adopter/nonadopter	Selected quotes
5	Nonadopter	I have no smartphone or computer.
9	Nonadopter	I have no phone or computer.
7	Nonadopter	Need an interpreter like you and pay for an interpreter's services.
11	Nonadopter	Need help with an interpreter.
4	Nonadopter	I survive from my social security check.
8	Nonadopter	Financial help to by phone and an interpreter.
3	Nonadopter	Education about telehealth and training about how to use it.
2	Nonadopter	I think training about telehealth and how to use it.
13	Nonadopter	Training about telehealth in the community or clinic.

Cultural Influence

Although the study participants were from the Asian American community in general, there were multiple ethnicities represented. One cultural influence identified by adopters was a preference for in-person care. However, even though adopters preferred in-person care, a recommendation by a doctor to use telehealth overrode the cultural-based preference of some adopters. It was clear that the convenience of using telehealth had more influence on their decision regarding whether to adopt telehealth. In contrast, other adopters retained their preference for in-person care despite the benefits of telehealth. Language was not an issue for most adopters, especially among second-generation Asian American individuals. Likewise, language was not an issue for Asian American individuals who spoke the same non-English language as their doctor. Still, some adopters did experience language barriers that affected their effective use of telehealth services.

Most nonadopters were not proficient in English, and this adversely affected their ability to use telehealth systems. Therefore, among participants who chose not to use telehealth, the primary concern was the language barrier; several nonadopters specifically cited the language barrier as prohibitive. A preference for in-person care was also cited as part of the reason participants did not adopt telehealth services.

Personal Innovativeness

Personal innovativeness is related to participants' comfort in using new technologies and their willingness to learn how to use telehealth [26,27]. All adopters were comfortable with telehealth technology. Participant occupations were often related to comfort level with telehealth systems. As many adopters were professionals in fields where the use of different technologies was required, they were easily able to attain proficiency on different platforms. Other adopters decided they were comfortable with new technologies because they were already using telehealth.

XSL•FO
RenderX

Among adopters, there was an overwhelming willingness to learn how to use telehealth for diabetes care. As all adopters except one were already using telehealth, they were willing to commit themselves to learning about the functionality and use of the system in even more detail.

In contrast, many nonadopters were not comfortable with new technologies due to legacy challenges, such as the language barrier. Some nonadopters who could not read or write in English would undoubtedly experience challenges using telehealth. However, regardless of language concerns, nonadopters were willing to learn how to use the technology.

Trust

Adopters must trust in the security, privacy, and confidentiality of their health information stored in health care systems to continue using telehealth services. Most adopters reported that they were comfortable with the security and privacy of their health information, expressing their confidence as being attributable to Health Insurance Portability and Accountability Act policies. Nonetheless, some adopters had concerns regarding security, as they reported worrying about the safety of their identifying information. In addition, some adopters were concerned that others would see their sensitive health information, although there were a few who felt that they could trust their physicians to keep their information safe and protected.

Most nonadopters did not trust telehealth systems. Their concerns were related to sharing their health information online and the security and privacy of information disclosed while using the systems. However, several nonadopters did not have an opinion one way or the other, as they had never used telehealth systems.

Behavioral Intention

Behavioral intention relates to the decision to continue or discontinue use of telehealth and the reasons for this decision [28]. All adopters except one said they would continue to use telehealth systems. The reasons that the majority of participants decided to continue using telehealth services were attributed to the benefits it provided. The one adopter who was noncommittal mentioned safety and security concerns about health information stored in the telehealth systems. Despite such concerns, most adopters were willing to continue using telehealth as a modality for managing diabetes.

Even though nonadopters had never used telehealth systems and despite the significant challenges identified, they were positive in their intentions toward the use of telehealth. A majority of nonadopters were willing to learn more about telehealth if someone taught them. In fact, nonadopters attributed their behavioral intentions to the need to improve their diabetes care. Other nonadopters cited the benefits of telehealth such as convenience or saving time and money as the reasons for their desire to learn. However, a few nonadopters refused to use telehealth in the future, citing existing language barriers.

Discussion

Summary of Key Findings

The analysis of the emergent themes from the qualitative data yielded 9 themes. The themes were actual use, effort expectancy, performance expectancy, social influence, facilitating conditions, cultural influence, personal innovativeness, trust, and behavioral intentions. Several subthemes also emerged from the themes. They included the possession of access devices, language barrier, the convenience and ease of using telehealth systems, cost and time effectiveness resulting from using telehealth systems, the role of physicians in recommending the use of telehealth systems, the need for training and education on the use of telehealth systems, insurance coverage of telehealth services, and an intention to continue using or start using telehealth systems.

Regarding actual use, the findings showed that the adopters who used telehealth systems cited the convenience, cost and time efficiency, and flexibility as some of their reasons for using telehealth systems, corresponding to Hu et al's [29] findings, showing that access to technology was a notable determinant in patients' interest in mobile health interventions. The nonadopters attributed their lack of awareness of the existence of telehealth systems and how to use them as some of the reasons for not using telehealth systems for the management of T2DM. Regarding effort expectancy, the findings showed that the participants perceived the use of telehealth systems to be easy. Concerning performance expectancy, telehealth systems improved convenience, enhanced access to health care services, were cost-efficient, saved time, and enabled flexibility in access to services. Regarding social influence, the findings showed that physicians, friends, and family played a significant role in promoting the adoption of telehealth. Similarly, Mora and Golden [30] found that family strongly influences individuals' diabetes management plans. Recommendations of telehealth from physicians, family, and friends were conspicuously absent among the nonadopters, while the adopters reported the influence of those recommendations on their decisions to adopt telehealth systems.

The facilitating conditions identified as being influential to the adoption of telehealth systems for the management of T2DM included access to devices, such as smartphones and computers. The coverage of telehealth services by insurance providers was another important facilitating condition. The other conditions were training on how to use telehealth systems and education on the use of telehealth systems for the management of T2DM. The findings also showed that cultural factors influenced the adoption of telehealth systems. The two most influential cultural factors were the language barrier and a preference for in-person care. Mora and Golden [30] also found that language barriers impacted diabetes management approaches.

Findings on personal innovativeness showed that there was an equal share of comfort and discomfort in using telehealth systems. The findings also showed that legacy challenges, such as the language barrier and lack of access to devices, affected the participants' ability to use telehealth systems. Regarding trust, participants shared concerns regarding the privacy,

XSL•FO
RenderX

security, safety, and confidentiality of the health information stored in telehealth systems. However, there was also trust in the professionalism of physicians and the safeguards provided by the Health Insurance Portability and Accountability Act policy. Concerning behavioral intentions, there was an overwhelming desire to continue using telehealth systems. Even those who had not used telehealth systems before were willing to use them, provided they were trained and offered financial aid through which to acquire access devices. The behavioral intentions of the participants were informed by the benefits of convenience, time savings, cost efficiency, and enhanced access to health care services.

The findings showed that cultural influences played a role in the participants' decision to adopt or not adopt telehealth systems for the management of T2DM. The most significant cultural factor influencing the decision not to use telehealth among the nonadopters was the language barrier. Research has shown that most of the telehealth platforms used in the United States use the English language [31]. The implications are that populations that are not fluent in English might find telehealth systems to be of limited benefit. This was certainly the case for many nonadopters. To address this challenge, the researcher recommends culturally-adapted telehealth systems that target underserved racial and ethnic minorities. Culturally adapted telehealth systems will not only address the language issue but also incorporate other cultural adaptations that would enhance the usability of telehealth systems for racial and ethnic minorities. Existing research supports this recommendation [32].

Barriers to Telehealth Adoption

The results of this qualitative case study indicate that language and cultural barriers significantly impact Asian American individuals' use of telehealth for managing T2DM. Language and cultural barriers have been recognized as key obstacles in health care access, often limiting patient engagement. For example, there is a limited availability of translated telehealth materials, and the scarcity of bilingual health care providers and interpreters makes it even more difficult for many groups to use this technology. When considering the culture of Asian American people, it also seems that there is a preference for in-person consultations with health care providers.

Another barrier to using telehealth is digital literacy and access, as many Asian American people do not have access to either technical support or patient education resources. Older adults especially seem to struggle with navigating telehealth platforms. Without proper training and support, these individuals remain excluded from telehealth-driven diabetes management.

The findings also highlight the limited recommendations of telehealth services by providers, which could be better promoted as a way for patients to improve their management of T2DM. Health care provider recommendations play a crucial role in shaping patient perceptions of telehealth. As the study uncovered, health care providers often do not actively recommend telehealth services, so many patients may not be aware of this option. Consequently, patients are also unaware of the many benefits of using telehealth services, which represents a significant barrier to its adoption. This highlights the need for provider engagement strategies to integrate telehealth into routine diabetes management.

Finally, technology and infrastructure disparities exacerbate other barriers to the use of telehealth services. Low-income individuals struggle with the cost of high-speed internet and smart devices, widening the gap of health care inequity [9]. For example, many patients, especially those with low income, are not able to obtain the devices (eg, smartphones and laptops) needed to access telehealth. Add internet connectivity issues and it is no surprise that many patients do not use telehealth. Addressing these disparities requires policy intervention that expands broadband access and subsidizes telehealth technology for underserved communities. Overall, the findings from this study align with existing literature on telehealth disparities among minority populations [11].

Limitations

Adopters were generally younger, more educated, and higher-income than nonadopters. These socioeconomic differences likely confound the observed adoption patterns. While our sample size precluded stratified or adjusted analyses, future studies should employ matched sampling or multivariable adjustment to disentangle cultural influences from socioeconomic status. Participants were recruited through two community-based health organizations in Missouri: A Federally Qualified Health Center and a local Asian-serving nonprofit clinic. Both provided limited interpreter support, which shaped recruitment feasibility and participant diversity.

The language barrier was a significant limitation during the collection of data. The participants were drawn from various ethnicities. Therefore, they had diverse native languages. Many were not proficient in English and did not share a common language with the researcher. The researcher relied on the services of an interpreter to translate the question to the participant and the response from the participant back to the researcher. The translation is prone to loss of meaning because the interpreter must interpret and decode the participant's words to derive their meaning. Context, whether personal or cultural, is important to the meaning of the participants' words. Furthermore, facial expressions, gestures, tone of voice, and pauses are also central to meaning. While the interpreter may translate the words said by the participant, an accurate translation of the meaning, semantics, and nuances behind the words may not always be possible. Therefore, some or the entire meaning of the communication may be lost during the translation.

The knowledge level of the nonadopters about telehealth systems is a limitation to the value of the data gathered from the cohort. Most of the participants were unaware of the existence of telehealth systems. Therefore, it is possible that they did not actually decide not to adopt telehealth for the management of their T2DM. The implications of this lack of awareness are that the information they provided may not have reflected the influence of the unique characteristics of telehealth on their nonadoption but rather an influence of their lack of awareness of the existence of telehealth systems and the value they provide. Lack of awareness of telehealth systems may explain responses to several prompts, most of which revolved around "I do not

XSL•FO
RenderX

know." The value of this information in answering the research questions was limited.

Conclusions

This qualitative case study identified unique characteristics, supported by the UTAUT model, that influenced the adoption intent and adoption of telehealth among Asian American people with T2DM in Missouri. Overall, there were many valuable insights into the cultural and technological barriers facing Asian American people when using telehealth.

Conflicts of Interest

None declared.

References

1. Gordon HS, Solanki P, Bokhour BG, Gopal RK. I'm not feeling like I'm part of the conversation: patients' perspectives on communicating in clinical video telehealth visits. J Gen Intern Med 2020 Jun;35(6):1751-1758. [doi: 10.1007/s11606-020-05673-w] [Medline: 32016705]
2. Sim R, Lee SWH. Patient preference and satisfaction with the use of telemedicine for glycemic control in patients with type 2 diabetes: a review. Patient Prefer Adherence 2021;15:283-298. [doi: 10.2147/PPA.S271449] [Medline: 33603347]
3. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
4. De Groot J, Wu D, Flynn D, Robertson D, Grant G, Sun J. Efficacy of telemedicine on glycaemic control in patients with type 2 diabetes: a meta-analysis. World J Diabetes 2021 Feb 15;12(2):170-197. [doi: 10.4239/wjd.v12.i2.170] [Medline: 33594336]
5. de Kreutzenberg SV. Telemedicine for the clinical management of diabetes; implications and considerations after COVID-19 experience. High Blood Press Cardiovasc Prev 2022 Jul;29(4):319-326. [doi: 10.1007/s40292-022-00524-7] [Medline: 35579849]
6. Gal RL, Cohen NJ, Kruger D, et al. Diabetes telehealth solutions: improving self-management through remote initiation of continuous glucose monitoring. J Endocr Soc 2020 Sep 1;4(9):bvaa076. [doi: 10.1210/jendso/bvaa076] [Medline: 32864542]
7. Ernawati U, Wihastuti TA, Utami YW. Effectiveness of diabetes self-management education (DSME) in type 2 diabetes mellitus (T2DM) patients: systematic literature review. J Public Health Res 2021 Apr 14;10(2):2240. [doi: 10.4081/jphr.2021.2240] [Medline: 33855427]
8. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
9. Diabetes and Asian Americans. US Department of Health and Human Services: Office of Minority Health. 2021. URL: https://minorityhealth.hhs.gov/diabetes-and-asian-americans [accessed 2026-01-13]
10. Anjana RM, Pradeepa R, Deepa M, et al. Acceptability and utilization of newer technologies and effects on glycemic control in type 2 diabetes: lessons learned from lockdown. Diabetes Technol Ther 2020 Jul;22(7):527-534. [doi: 10.1089/dia.2020.0240] [Medline: 32522031]
11. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
12. Goh KLS, Lee CS, Koh CHG, et al. Evaluating the effectiveness and utility of a novel culturally-adapted telemonitoring system in improving the glycaemic control of Asians with type-2 diabetes mellitus: a mixed method study protocol. Trials 2021 Apr 26;22(1):305. [doi: 10.1186/s13063-021-05240-6] [Medline: 33902656]
13. Khalid A, Dong Q, Chuluunbaatar E, Haldane V, Durrani H, Wei X. Implementation science perspectives on implementing telemedicine interventions for hypertension or diabetes management: scoping review. J Med Internet Res 2023 Mar 14;25:e42134. [doi: 10.2196/42134] [Medline: 36917174]
14. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice, 4th edition: Sage Publications; 2015. URL: https://discovered.ed.ac.uk/discovery/fulldisplay?vid=44UOE_INST:44UOE_VU2&mapsearch.scope=UoE&mapdb=Everything&mapdocid=alma9923876733302468&mapdlang=en&mapcontext=L&mapisFrb=true [accessed 2026-01-13]
15. American Community Survey data. US Census Bureau. 2022. URL: https://www.census.gov/programs-surveys/acs/data.html [accessed 2026-01-21]
16. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1. [doi: 10.11648/j.ajtas.20160501.11]
17. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019;7(2):e000057. [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
18. Guest G, MacQueen K, Namey E. Applied Thematic Analysis: Sage Publications; 2012. [doi: 10.4135/9781483384436]

XSL•FO
RenderX

19. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Dec 1;16(1). [doi: 10.1177/1609406917733847]
20. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
21. Suki NM, Suki NM. Determining students' behavioural intention to use animation and storytelling applying the UTAUT model: the moderating roles of gender and experience level. Int J Manag Ed 2017 Nov;15(3):528-538. [doi: 10.1016/j.ijme.2017.10.002]
22. Truong M, Yeganeh L, Cook O, Crawford K, Wong P, Allen J. Using telehealth consultations for healthcare provision to patients from non-Indigenous racial/ethnic minorities: a systematic review. J Am Med Inform Assoc 2022 Apr 13;29(5):970-982. [doi: 10.1093/jamia/ocac015] [Medline: 35150266]
23. Batucan GB, Gonzales GG, Balbuena MG, Pasaol KRB, Seno DN, Gonzales RR. An extended UTAUT model to explain factors affecting online learning system amidst COVID-19 pandemic: the case of a developing economy. Front Artif Intell 2022;5:768831. [doi: 10.3389/frai.2022.768831] [Medline: 35573898]
24. Chao CM. Factors determining the behavioral intention to use mobile learning: an application and extension of the UTAUT model. Front Psychol 2019;10:1652. [doi: 10.3389/fpsyg.2019.01652] [Medline: 31379679]
25. Venkatesh V, Sykes TA, Zhang X. "Just what the doctor ordered": a revised UTAUT for EMR system adoption and use by doctors. Presented at: 44th Hawaii International Conference on System Sciences (HICSS 2011); Jan 4-7, 2011. [doi: 10.1109/HICSS.2011.1]
26. Bandyopadhyay K, Fraccastoro KA. The effect of culture on user acceptance of information technology. Commun Assoc Inf Syst 2007;19(19):23-29. [doi: 10.17705/1CAIS.01923]
27. Oh JC, Yoon SJ. Predicting the use of online information services based on a modified UTAUT model. Behav Inf Technol 2014 Jul 3;33(7):716-729. [doi: 10.1080/0144929X.2013.872187]
28. Twum KK, Ofori D, Keney G, Korang-Yeboah B. Using the UTAUT, personal innovativeness and perceived financial cost to examine student's intention to use e-learning. J Sci Technol Policy Manag 2022 Aug 3;13(3):713-737. [doi: 10.1108/JSTPM-12-2020-0168]
29. Hu L, Wyatt LC, Mohsin F, et al. Characterizing technology use and preferences for health communication in South Asian immigrants with prediabetes or diabetes: cross-sectional descriptive study. JMIR Form Res 2024 Apr 26;8:e52687. [doi: 10.2196/52687] [Medline: 38669062]
30. Mora N, Golden SH. Understanding cultural influences on dietary habits in Asian, Middle Eastern, and Latino patients with type 2 diabetes: a review of current literature and future directions. Curr Diab Rep 2017 Oct 23;17(12):126. [doi: 10.1007/s11892-017-0952-6] [Medline: 29063419]
31. Payán DD, Frehn JL, Garcia L, Tierney AA, Rodriguez HP. Telemedicine implementation and use in community health centers during COVID-19: clinic personnel and patient perspectives. SSM Qual Res Health 2022 Dec;2:100054. [doi: 10.1016/j.ssmqr.2022.100054] [Medline: 35169769]
32. Litchman ML, Kwan BM, Zittleman L, et al. A telehealth diabetes intervention for rural populations: protocol for a randomized controlled trial. JMIR Res Protoc 2022 Jun 14;11(6):e34255. [doi: 10.2196/34255] [Medline: 35700026]

Abbreviations

T2DM: type 2 diabetes mellitus
UTAUT: unified theory of acceptance and use of technology

Edited by S Li; submitted 08.Apr.2025; peer-reviewed by J Lin, Y Song; revised version received 04.Dec.2025; accepted 17.Dec.2025; published 03.Feb.2026.

Please cite as:
States DG
Cultural and Technological Barriers to Telehealth Adoption for Type 2 Diabetes Management Among Asian American Patients: Qualitative Case Study
JMIR Diabetes 2026;11:e75689
URL: https://diabetes.jmir.org/2026/1/e75689
doi:10.2196/75689

© Devi Gurung States. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 3.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR

XSL•FO
RenderX

Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Process of Developing an Intervention to Increase Awareness of Cardiovascular Risk for Persons With Type 2 Diabetes: Co-Creation Study

Anna-Lena Stenlund[1], RN, MSc; Karin Hellström Ängerud[1], RN, PhD; Mikael Lilja[2], MD, PhD; Julia Otten[2], MD, PhD; Lena Jutterström[1,2], RN, PhD

[1]Department of Nursing, Umeå University, Campus Skellefteå, Bockholmsvägen 23, Umeå, Skellefteå, Sweden
[2]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

Corresponding Author:
Anna-Lena Stenlund, RN, MSc
Department of Nursing, Umeå University, Campus Skellefteå, Bockholmsvägen 23, Umeå, Skellefteå, Sweden

Abstract

Background: Many persons with type 2 diabetes (T2D) lack risk awareness or underestimate their cardiovascular risk. Although health care professionals in primary health care strive to implement risk-awareness strategies for cardiovascular risk, persons with T2D report a lack of meaningful dialogue with health care professionals. Co-creation is grounded in participatory action research and involves participants as equal partners across all stages of a project. This study describes the development of an intervention to increase cardiovascular risk awareness in people with T2D.

Objective: This study aims to describe the co-creation process of developing an intervention to increase awareness of cardiovascular risk in persons with T2D.

Methods: A co-creative design was used to develop an intervention following a participatory action research framework. Four workshops with persons with T2D, diabetes specialist nurses, and physicians in primary health care explored communication about cardiovascular risk, co-identified needs, co-designed solutions, tested prototypes, and redefined and retested the content of the intervention. The data were analyzed using reflexive thematic analysis.

Results: The analysis identified 4 themes: co-define: taking the person's voice into account; co-design: problem-solving and generating ideas; prototype and test: drafting intervention proposals; and redefine and retest: reviewing suggested interventions. The workshop discussions highlighted the need for new interventions, including a risk assessment tool, a patient handbook, material to prompt reflection, and a web education for specialist diabetes nurses.

Conclusions: This study demonstrates the value of co-creation, which was used to develop an intervention to enhance cardiovascular risk awareness in persons with T2D. Diabetes specialist nurses need to explore patients' perceptions of risk and provide space for emotional responses. The web education is intended to strengthen the person-centered approach of diabetes specialist nurses, the patient handbook encourages reflection and dialogue on personal risk, and the risk assessment tool visualizes individual risk. These components may contribute to increased awareness of cardiovascular risk.

(JMIR Diabetes 2026;11:e85748) doi:10.2196/85748

KEYWORDS

co-creation; participatory action research; PAR; intervention development; type 2 diabetes; cardiovascular risk: risk awareness

Introduction

Cardiovascular Risk in Type 2 Diabetes

Persons with type 2 diabetes (T2D) face an elevated risk of micro- and macrovascular complications, including a two- to four-times increased likelihood of developing cardiovascular diseases such as myocardial infarction and stroke [1]. Early detection of T2D and timely optimization of hemoglobin A1c (HbA_{1c}) and other risk factors are crucial for preventing cardiovascular events [2]. Despite this, not all persons with T2D reach treatment goals [3]. The prevention of cardiovascular risk involves addressing behavioral risk factors, such as regular exercise, healthy eating, smoking cessation, and medical treatment [4]; effective risk communication reduces the overall risk and enhances persons' self-perceived risk motivation [5]. Health care professionals (HCPs) in primary health care are expected to provide person-centered communication of cardiovascular risk; however, research shows that risk communication is often one-sided, with HCPs mainly informing rather than engaging with persons [6].

XSL•FO
RenderX

Background

Many persons with T2D lack risk awareness or underestimate their cardiovascular risk, particularly in the absence of symptoms [7]. Although the link between T2D and cardiovascular risk is well established, effective risk communication regarding myocardial infarction and stroke remains a challenge [8]. Person-centered care can improve risk awareness and self-management. Interventions building on the self-determination theory also improve health behavior for self-management and motivation [6]. Interventions based on self-determination theory can improve changes in health behavior and autonomy motivation for self-management (ie, 9) and can significantly reduce cardiovascular risk [9]. However, HCPs find it difficult to perform person-centered care in clinical practice [10].

HCPs in primary health care should strive to implement risk-awareness strategies for cardiovascular risk in T2D; however, persons with T2D report a lack of meaningful dialogue with HCPs [8] and express a desire for more personalized, participatory communication. Risk communication must be individual-specific enough to make a person with T2D perceive their own risk without inducing fear or anxiety but rather fostering understanding and motivation [11]. Trust and shared decision-making have been highlighted as important to effective risk communication and patient motivation [12]. While HCPs are expected to communicate cardiovascular risk to persons with T2D [7], there is limited guidance on how to do so effectively; therefore, more research in this area is needed.

Co-creation has emerged as a promising approach to intervention development and increasing participation and involvement in health care [13]. Co-creation is rooted in participatory action research (PAR) and involves participants as equal partners throughout all stages of a project [14]. It facilitates joint problem identification, shared understanding, and context-specific solutions [15] and enhances the effective development of interventions [14]. Co-creation design uses methods and strategies to identify problems, improve user experience, and adapt functionality [15]. Studies within PAR have shown that when participants are involved in shaping the design of an intervention, they are more likely to adopt it in clinical practice [16].

Effective interventions should enhance risk awareness, strengthen motivation, and support behavior change related to elevated cardiovascular risk. This requires incorporating diverse perspectives and ensuring that interventions are both theoretically sound and practically applicable [14]. Interventions that raise awareness of cardiovascular complications, while also addressing attitudes and self-management, benefit not only persons with T2D but also the health care system [13].

Our previous qualitative studies involving persons with T2D and HCPs in primary health care indicate that cardiovascular risks are not communicated sufficiently and that the risk communication that does occur during clinical consultations is often ineffective for persons with T2D [17]. There is a need for more interventions developed using co-creation to improve risk communication and thereby enhance risk awareness among persons with T2D. This study aims to describe the process of developing an intervention intended to strengthen the awareness of persons with T2D of their cardiovascular risk and foster more effective self-management of the condition.

Aim of the Study

This study aims to describe the co-creation process of developing an intervention with the aim of increasing awareness of cardiovascular risk in persons with T2D.

Methods

Design

A co-creation study design within a PAR framework [13,14] was used.

Study Setting and Recruitment

This study was carried out as a series of workshops that followed a PAR framework and involved persons with T2D and HCPs working in primary health care in northern Sweden. To recruit participants, posters were distributed to primary health care units. No one answered; therefore, diabetes specialist nurses were contacted in order to suggest participants in the form of both persons with T2D and HCPs. Thereafter, a snowball sampling was conducted for persons with T2D, diabetes specialist nurses, and physicians who fulfilled the inclusion criteria. The inclusion criteria for persons with T2D were a diagnosis of T2D, being 18 years or older, and being Swedish-speaking. The HCPs needed to be diabetes specialist nurses or physicians working with diabetes in primary health care.

In total, 10 persons with T2D (median age 68.5 y, range 52 - 79) and 5 HCPs (median age 53 y, range 32 - 59) participated (Table 1). The HCPs were 3 diabetes specialist nurses and 2 physicians, with working experience in primary health care ranging from 2 to 20 years (Table 2). Before the workshops, the participants received information about the study both orally and in writing and had the opportunity to ask questions. Written consent was obtained from all the participants. Workshops 1 and 2 included persons living with T2D and 2 researchers. Workshops 3 and 4 included diabetes specialist nurses, physicians, and 2 researchers. Thus, a different set of participants attended each workshop.

XSL•FO
RenderX

Table . Demographic characteristics of persons with type 2 diabetes (n=10).

Characteristics	Values
Diabetes duration (y), median (IQR)	13.5 (2 - 21)
Age (y), median (IQR)	68.5 (52 - 79)
Gender, n	
Female	5
Male	5
Education level	
Primary school level or upper secondary school level	6
University or higher education level	4
Cardiovascular history, n	
Yes	2
No	8
Treatment for high blood pressure or blood lipids, n	
Yes	8
No	2
Smoking, n	
Yes	0
No	10

Table . Demographic characteristics of the health care professionals (n=5).

Characteristic	Participants
Diabetes specialist nurses	3
Physicians	2
Professional experience (y), range	2 - 20
Age (y), median (range)	53 (32 - 59)
Gender, n	
Female	4
Male	1
Digital literacy, n	
Yes	5
No	0

Data Collection

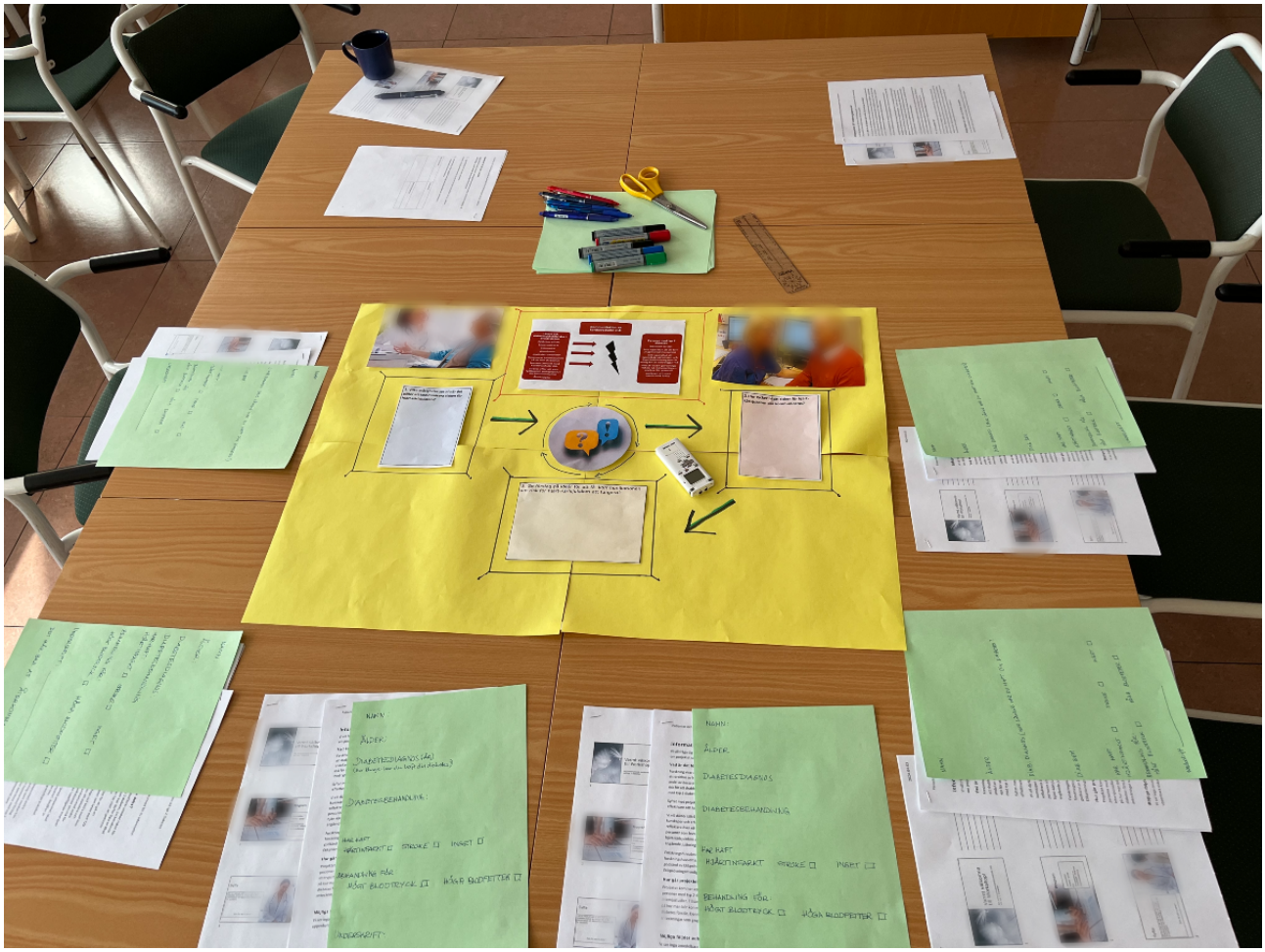
Overview

Before the workshops, the researchers conducted a review of the results of both their own previous studies and those of other researchers. They planned and designed the structure and flow of the sessions and prepared probes, toolkits, the agenda, activities, and supporting materials. The workshop materials included presentations, images, worksheets, and scenario-based exercises. Figure 1 shows examples of probes and toolkits for the participants of workshops 1 and 2. In workshops 3 and 4, PowerPoint material was used to share the findings from previous workshops.

Data collection was conducted at the workshops involving persons with T2D and 2 researchers (workshop 1, n=8; workshop 2, n=6) and the HCPs and 2 researchers (workshop 3, n=4; workshop 4, n=5). The workshop discussions included 2 researchers, who facilitated the workshops and produced field notes. The workshops were held in Swedish, audio-recorded, and transcribed verbatim and analyzed using reflexive thematic analysis. Each workshop lasted between 90 and 120 minutes. Workshops 1 and 2 were held in person at the university; workshops 3 and 4 were held online. All workshops were held between May and November 2024.

XSL•FO
RenderX

Figure 1. Examples of probes and toolkits for the participants of workshops 1 and 2.



The Co-Creation Process Within a PAR Framework

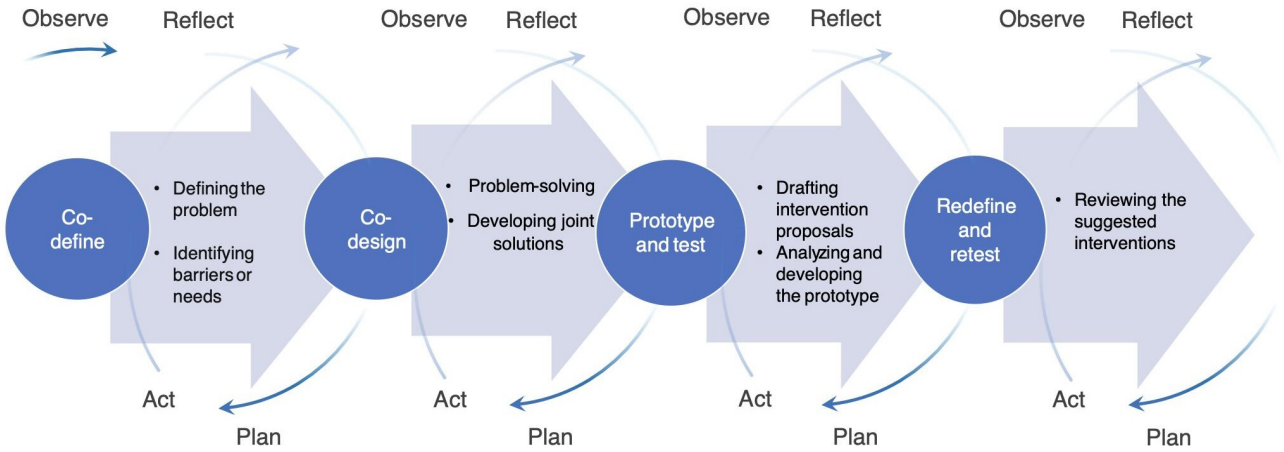
PAR is a qualitative methodology where participants actively engage in the research process [18]. It values the knowledge and experiences of participants, makes use of democratic processes for positive change, and emphasizes action [19]. The process involves iterative cycles of relationship-building and colearning to identify problems, design solutions, and reflect on actions over the course of at least three cycles [20]. Each cycle consists of 4 stages: observe, reflect, plan, and act [16]. The observation stage involves clarifying and identifying problems, needs, and barriers. In the reflection stage, the participants analyze and reflect, listen to different perspectives, and share ideas and experiences. The planning stage involves clarifying and drafting proposals and developing an action plan. The action stage involves visualizing the ideas, shaping and revising the intervention [16]. One review found that PAR is frequently applied in several phases of the research process such as data collection, research design, and recruitment and is grounded in the real-world experiences of participants [18]. In this study, 4 cycles were conducted. Persons with T2D and HCPs were actively involved in data collection, interpretation of findings, co-defining, co-designing, prototyping, and testing, as well as redefining and retesting the intervention.

Figure 2 illustrates the co-creation process of the PAR framework followed in this study: co-define, co-design, prototype and test, and redefine and retest, influenced by previous studies [16,19-21]. The co-define stage included the identification of barriers and needs and problem formulation. The co-design stage included solution development, and the prototyping and testing stage involved evaluating and forming intervention prototypes. The redefine and retest stages included improvement and evaluation and revising the intervention (Figure 2) [16].

The facilitators (researchers) in this study did not propose ideas; instead, they encouraged the participants to share perspectives, suggest strategies, and co-create solutions. Based on these discussions, the researchers developed materials to facilitate risk communication and strengthen risk awareness among persons with T2D.

XSL•FO
RenderX

Figure 2. An overview of how the participatory action research (PAR) process proceeded and how the results evolved progressively over the course of the 4 co-creation cycles. The figure is influenced by previous research [16,19-21].

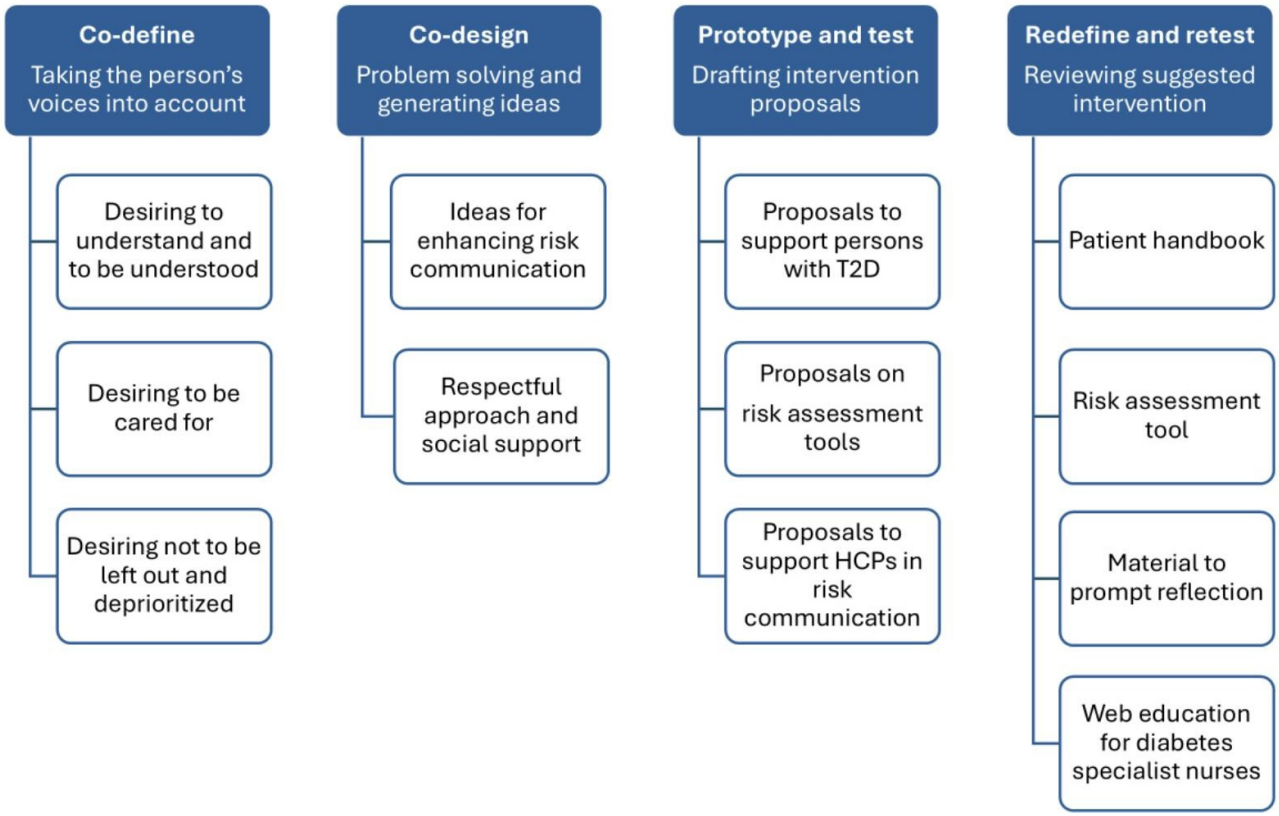


Data Analysis

Reflexive thematic analysis was conducted to analyze the workshop transcripts, following the steps outlined by Braun and Clarke [22]. To become *familiar with the data*, the first author read the verbatim transcripts (58,596 words) and field notes multiple times, making notes and marking sentences and paragraphs that were relevant to the aim of the study [22]. The initial coding was carried out inductively by the first author using MAXQDA 2022 to facilitate the organization, sorting, and coding of the data. A manifest analysis approach was used, which involved highlighting sections of text, interpreting the obvious in the text, identifying relevant phrases and sentences, developing codes (1123 codes) to describe their content, and sorting the codes into groups. Paraphrasing and marginal notes were incorporated to preserve contextual relevance and support the analytical process when developing subthemes. Thereafter, the first author incorporated latent thematic analysis to explore the deeper meanings of the text. In this study, guided by the principles of PAR, the analytical process shifted from an inductive to a deductive approach when *generating themes*. Themes were defined through iterative collaborative reflection (ALS, LJ), in which the researchers identified patterns in the material and grouped them into themes and subthemes based on the content (Figure 3). The validity and reliability of the themes were ensured through the development of a thematic map that illustrated the relationships among codes, themes, and subthemes (Multimedia Appendix 1). The themes were then defined and named through iterative discussions between the 3 authors until consensus was reached. Subsequently, the entire research group met to agree on the final themes (ALS, LJ, KHÄ, JO, ML) (Multimedia Appendix 1).

XSL•FO
RenderX

Figure 3. Overview of the themes and subthemes. HCP: health care professional; T2D: type 2 diabetes.



Ethical Considerations

Ethical approval was granted by the Swedish Ethical Review Authority (2018-12-11; Dnr 2023-06037-02). The clinical director and department heads of the primary health care centers approved the study. The participants were informed verbally and in writing about the study's aim, their participation and right to withdraw, and confidentiality. Written informed consent was obtained from all participants and they were informed that participation was voluntary and that they could withdraw from the study at any time. In addition, informed consent was obtained from the participants to use their images for publication purposes. Furthermore, they were guaranteed confidentiality, with the results presented at group level to avoid identification and quotations deidentified. The participants did not receive compensation.

Rigor and Reflexivity

The reflexive thematic analysis reporting guidelines were used to ensure the rigor, coherence, and reflexive openness of the study (Checklist 1) [23]. The co-creation design ensured credibility [24]. The research process was transparently documented, with data collection, analysis, and interpretation detailed. Trustworthiness was enhanced by clearly describing the methodology and maintaining reflexivity to minimize bias. The researchers' nursing and teaching experiences enabled them to ask relevant questions and helped build trust with the participants. To mitigate bias, the researchers adopted a reflexive stance, systematically examining their assumptions and preconceptions. This reflexivity was essential to uphold ethical rigor and enhance the trustworthiness of the data collection process. The researchers came from diverse backgrounds and held regular meetings to critically reflect on the data. Confirmability was maintained through a systematic coding process and rigorous data analysis, which ensured that the findings were grounded in the accounts of the participants. Credibility was strengthened by building trust, clarifying goals, and fostering inclusive dialogue with the participants. Transferability was supported by thorough documentation of context, including participant demographics (eg, age, disease duration, previous cardiovascular events) and setting.

Results

Themes and Subthemes

The analysis resulted in 4 themes and 12 subthemes, wherein the participants identified challenges, generated ideas, prototyped intervention options, and evaluated these options for cardiovascular risk communication (Figure 3). The co-creation process and the findings of the thematic analysis are presented under 4 identified themes: co-define: taking the person's voice into account; co-design: problem-solving and generating ideas; prototype and test: drafting intervention proposals; and redefine and retest: reviewing suggested interventions.

Co-Define: Taking the Person's Voice Into Account

Overview

In cycle 1, the focus was on identifying communication barriers and needs from the perspectives of the persons with T2D. The participants' actions included discussing problems, challenges, and needs and sharing experiences. The researchers began every workshop by introducing the topic and purpose of that

XSL•FO
RenderX

workshop, as well as discussing previous research and workshops. The participants were encouraged to discuss the findings of previous research, including our previous interview studies with persons with T2D and diabetes nurses and physicians.

Results of the Thematic Analysis

Desiring to Understand and to Be Understood

Persons with T2D want HCPs to communicate cardiovascular risk in various ways, and in a manner that they can understand, to ensure clarity and comprehension. The participants stated that medical jargon and complex terminology should be avoided, and clear and simple language should be prioritized:

> *Don't say cardiovascular risk; say myocardial infarction and stroke.* [Participant, cycle 1]

The participants described not fully understanding their risks of myocardial infarction and stroke, and many claimed that they had never heard that T2D could lead to myocardial infarction and stroke. The participants expressed a desire for clarity regarding the meaning of test results, such as explaining HbA_{1c} and low-density lipoprotein cholesterol values, and how these impact cardiovascular risk.

> *Just this thing about what could happen if I've had this for so long – am I having a myocardial infarction? Is a stroke coming? But no one has told me about that.* [Participant, cycle 2]

The participants shared the belief that HCPs were too vague when discussing these risks, and the results indicate that the participants believe that HCPs avoid discussing risks and even downplay them because they are afraid of scaring patients or handling emotional reactions.

> *Yes, and I also think it's too sensitive for many people, even within healthcare, to talk about this. Yet, it's something we all know – that we're all going to die someday.* [Participant, cycle 1]

Many of the participants expressed a preference for receiving guidance in both written and verbal form, to reinforce their understanding of their cardiovascular risk and help them with the self-management activities involved in reducing it. The participants expressed the need for simple, relevant materials and guidance regarding where they could independently explore information suited to their individual needs.

> *Sit and listen first, but then I would like to have it written down on paper – yes, almost. I mean, clear communication that I can go back to and read later.* [Participant, cycle 1]

Desiring to Be Cared For

The participants described the importance of their interactions with HCPs, highlighting these encounters as a crucial aspect of their diabetes care. However, a recurring concern was the perceived lack of curiosity, interest, and engagement from HCPs. The persons with T2D stated that HCPs should show more interest in them as persons and ask more questions to help them feel more involved in their own care. The participants highlighted the need for undivided attention, emotional presence on the part of HCPs, and meaningful dialogue, as well as a communication style that conveys hope, encouragement, and care without blame. Persons with T2D described experiences of inconsistent communication and perceived nonchalance, which contributed to feelings of being dismissed or undervalued in the health care encounter. The participants described wanting to feel truly heard and cared for:

> *In that moment, I want to feel like you care only about me and that she should go through what has happened previously – what my blood sugar levels were like back then – and also explain the results of the tests I took a few days earlier. And maybe answer questions like: how does this work? Is it good or bad, or somewhere in between? And what can I do?* [Participant, cycle 2]

The results show that persons with T2D would like to feel trust and understanding during meetings with HCPs but are often not even asked how they are feeling:

> *I have never experienced a diabetes nurse asking anything like, how are you feeling? Are you affected by seasonal depression? I mean, she could ask anything … but never.* [Participant, cycle 1]

Persons with T2D expressed that stress and the emotional burden of managing diabetes further complicate risk communication, noting that these factors can reduce their attention and contribute to feeling overwhelmed. Furthermore, the results revealed that this emotional topic was seldom addressed in risk communication and that HCPs tended to follow their own agendas.

Desiring Not to Be Left Out and Deprioritized

Insufficient follow-ups, limited continuity of care with HCPs, time constraints in health care, and a sense that patients were merely being "checked off" rather than supported were described. The results indicate that HCPs control and determine the content of discussions during diabetes appointments, contributing to persons with T2D feeling excluded and perceiving that decisions are made over their heads. This perceived nonchalance reinforced feelings of being dismissed or undervalued during health care encounters. As one participant expressed:

> *Just for a moment, and the moment is decided by her, how much time she has, and then we don't talk about anything else. We switch to the "red light" at the same time; it feels like you said – I'm here to find out something. I go there with the suspicion that maybe I'll die tomorrow, or how else should I feel? I don't know anything else.* [Participant, cycle 2]

The results illustrate that persons with T2D experience a lack of structure in health care and feel that they are not prioritized when their needs are not addressed. The participants described negative attitudes and stigma relating to diabetes on the part of HCPs, and these influence their health care experiences and further contribute to their sense of being part of a deprioritized group.

XSL•FO
RenderX

> *It's important not to feel that the T2D patient group is deprioritized or placed after all other patient groups.* [Participant, cycle 2]

Persons with T2D described feeling that they must take responsibility for their care on their own. Several participants expressed dissatisfaction with having to remind HCPs about the regular collection of blood samples:

> *There is a lot of engagement required from the patient. You have to remind them about testing and contact with the diabetes nurse varies. Sometimes, you have to beg to get tests done, and most of the time, you have to find out about risks yourself. They don't ask about family history.* [Participant, cycle 2]

The results revealed that persons with T2D also want diabetes specialist nurses to acknowledge the emotional aspects during consultations.

Co-Design: Problem-Solving and Generating Ideas

Overview

The co-design stage involved the development of ideas and joint solutions, which were based on previous studies and the results of the first cycle of inquiry. Several suggestions for improving risk awareness in T2D were presented. The findings indicate that effective risk communication should be individualized, in verbal and written form, be direct rather than vague, be accessible, and easy for persons with T2D to refer to.

Results of the Thematic Analysis

Ideas for Enhancing Risk Communication

To address barriers to risk communication, both persons with T2D and HCPs highlighted the need to acknowledge individual perspectives on risk. The results highlighted the importance of combining verbal and written communication, such as note-taking and providing clear summaries, to support the understanding of patients. The results of cycle 3 suggest that HCPs see the patient group as diverse and recognize that individuals require different approaches that are adapted based on preferences and needs. Relying solely on digital communication is not suitable, as many people with T2D are older and not digitally literate. According to the results, using paper-based documents is a helpful option.

One suggestion for improving risk communication was to encourage individuals with T2D to reflect on and articulate their perceptions of their own cardiovascular risk. Questions that prompt reflection were considered to be essential, including those that explore patients' understanding of cardiovascular risk, self-management activities, emotional responses, and future expectations, as well as the role of family support.

In cycle 3, a risk assessment tool was proposed to support follow-ups and help individuals with T2D to monitor the progression of their cardiovascular risk over time. The participants emphasized the importance of providing patients with laboratory values related to their own risk, along with asking them exploratory questions to explore how they perceive that risk.

An idea that emerged involved illustrating cardiovascular risk levels using colors to enhance understanding. One suggestion was to develop a patient handbook that includes essential information and material that is intended to prompt reflection, to improve the person's risk awareness.

> *A kind of book that the patient takes with them. Maybe they would have tasks in the book, such as: find out this for the next time you meet with your diabetes nurse or doctor. There are things you need to understand and know, and that's what we should focus on.* [Participant, cycle 3]

Respectful Approach and Social Support

The results showed the importance of social support and a sense of equality during health care interactions. The HCPs asserted that building trust, fostering engagement, and encouraging change were essential for persons with T2D to feel comfortable opening up and sharing their thoughts and feelings about their risk. The results showed that the HCPs feel that it is important to be professional, with the aim of reducing inequality in interactions. This involves not underestimating individuals with T2D, treating them with respect, instilling hope, allowing questions, avoiding blame, and demonstrating genuine care and trustworthiness. Addressing power imbalances was posited as an aspect to improve the interaction between persons with T2D and HCPs. Social support and peer interaction were also highlighted as valuable components, with suggestions of group meetings and family-inclusive consultations to improve risk awareness. Some HCPs argued that a relative accompanying a person with T2D to an appointment meant that information could be discussed further at home:

> *If you [the patient] have the opportunity to have someone from your close family or a partner with you, who will take notes, it's really helpful.* [Participant, cycle 3]

Prototype and Test: Drafting Intervention Proposals

Overview

At the beginning of cycle 3, the researchers presented previous discussions, ideas, and drafts from the co-define and co-design stages. The HCPs evaluated the discussions and scenarios used during workshops 1 and 2 with persons with T2D by analyzing, reflecting on, and discussing difficulties and strategies relating to, and the structure of interventions. The HCPs were encouraged to share the challenges they face in communicating cardiovascular risk, propose strategies, and develop the design and structure of the prototype intervention. The workshops were structured to encourage collaboration and facilitate the gathering of valuable insights regarding the communication of cardiovascular risk.

Results of the Thematic Analysis

The proposals identified as important elements of an intervention to increase risk awareness are described below.

Proposals to Support Persons With T2D

Based on the results of cycle 3, ideas emerged regarding a patient handbook to support persons with T2D. This was

XSL•FO
RenderX

intended to help them reflect on and understand their risk of myocardial infarction and stroke individually, with family members, and in dialogue with diabetes specialist nurses and physicians, emphasizing person-centered care.

> *Having a small book, where you write down your questions, so you can bring them to the meeting we're going to have, because you'll never remember everything. We'll talk about so many things, and then you'll remember the things you really want to know.* [Participant, cycle 3]

Another suggestion was to develop discussion materials that could facilitate risk communication and encourage a progression in the discussions – moving forward, rather than repeating the same information. It was expressed that diabetes specialist nurses and physicians need to promote patient involvement, actively listen to the issues or topics that persons with T2D wish to discuss, and demonstrate genuine interest in the patient, rather than following their own agenda.

> *And then the patient can go back and say: oh, right, we talked about this last time. What was it again? And then there could be some sort of reference so that as a patient, you can go in here and read a bit more about it. It doesn't need to be anything overly complicated.* [Participant, cycle 3]

The results highlighted the use of materials to provoke reflection, to help persons with T2D assess their level of cardiovascular risk. Such materials were felt to help the persons to reflect independently, with family members, or in collaboration with health care professionals, all of which serve to enhance their understanding of the risks involved.

> *One should be able to include questions that the patient, on their own or together with someone, can reflect on.* [Participant, cycle 3]

At the same time, it was noted that a book may not provide suitable support to all persons with T2D. One physician noted the importance of integrating the personal handbook into visits, but that not all persons may be ready or interested in engaging with such exercises.

Proposals on Risk Assessment Tool

During cycle 3, the participants suggested using visual tools, such as color-coded images to provide information regarding different levels of risk for myocardial infarction and stroke (low, medium, high). However, some of the participants described red as an "angry" color, prompting discussions about how visual pathways can be used to illustrate steps toward lower risk levels and improved motivation for self-management activities.

> *It should visually show something like, you're at high risk, but the red feels a bit angry to me. Could we create pathways forwards from the red? I mean, this might sound childish, but maybe we could make images that show how to move on from there …* [Participant, cycle 3]

Proposals to Support HCPs in Risk Communication

One suggestion that emerged from the co-creation process was that HCPs need to improve their ability to adopt a person-centered approach when communicating cardiovascular risk. Therefore, a web education was developed that addresses risk communication. Intervention ideas from the co-creation process were used in the development of the web-based education, which was intended for diabetes specialist nurses and focuses on important aspects of person-centered risk communication and exploring individual risk. The web-based education explains the aim of the patient handbook and how to use risk assessment tools and reflective questions.

> *The networking sessions we have each semester are more about receiving communication on current topics, like something about test strips, procurements, and such. Perhaps what we need is more about how to use tools and exchange ideas on how to work, how to communicate, and how to reach these persons. What am I doing wrong?* [Participant, cycle 3]

Redefine and Retest: Reviewing Suggested Interventions

Overview

The researchers analyzed results from previous workshops and drafted intervention proposals for the participants to review during the redefine and retest stages. The Consolidated Framework for Implementation Research (CFIR) was used to develop the intervention, considering potential barriers and facilitators in risk communication (Multimedia Appendix 2). We also considered the nature of the intervention before presenting the proposal to HCPs in workshop 4 (Multimedia Appendix 1). The HCPs were encouraged to evaluate strategies and tools and suggest improvements. The proposed intervention focused on a patient handbook for persons with T2D, a risk assessment tool, to prompt reflection, and a web education to support diabetes specialist nurses. The workshop addressed how these tools could support the communication of cardiovascular risk and what additional support might be needed. Insights from this stage guided the refinement of the intervention's content and scope.

After the development of the preliminary intervention, the researchers collected feedback from persons with T2D. They were asked to evaluate the personal handbook, the risk assessment tool, reflective material, and the web education for diabetes specialist nurses. The intervention materials were distributed to the participants, who were invited to review the content and provide feedback on its relevance, clarity, and usability. Feedback was collected either in writing via email or verbally during scheduled follow-up conversations.

Results of the Thematic Analysis

Patient Handbook

The handbook allows patients to reflect on their risk, consider strategies for managing it, and set goals to promote engagement and risk awareness. It was designed by the researchers to provide concise information on cardiovascular risk factors and practical guidance for reducing cardiovascular risk, as well as questions to prompt reflection to facilitate exploration of personal understanding of cardiovascular risk.

XSL•FO
RenderX

In workshop 4, the HCPs stated that opportunities for reflection and goal-setting could be valuable in enhancing the understanding and awareness of the risks associated with T2D for patients.

> *Encouraging persons to reflect on and write down their thoughts and goals can strengthen awareness of the risk of myocardial infarction and stroke.* [Participant, cycle 4]

Feedback on the patient handbook during workshop 4 indicated that it can serve as a valuable tool for reflection for people with T2D, facilitate conversations with relatives and friends, and support collaboration with diabetes specialist nurses and physicians.

> *It's helpful to raise questions about what they believe regarding their abilities and how much confidence they have in themselves, so they can discuss it during the diabetes appointment.* [Participant, cycle four]

The results indicated that the patient handbook needed to be designed to encourage reflection and active participation on the part of the patient. The participants noted that, unlike traditional diabetes visits, the handbook provides an opportunity for them to consider their own perspectives and generate discussion during consultations, increasing their involvement:

> *No, the persons, those who receive this in their hands, of course, they can have their opinions, but it's interesting and refreshing that they get to reflect in a way they didn't before. We used to have these health declaration forms that they received at home before visits, and they would fill them out. This is like a different kind of thought-provoker, I think, and it could probably raise some questions that we can discuss during the visits.* [Participant, cycle 4]

Risk Assessment Tool

The HCPs expressed a need for a structured overview of the results of laboratory tests, along with other health parameters that are central to assessing and managing T2D. A risk assessment tool was developed by the researchers to visualize individual risk factors based on HbA_{1c}, blood lipids, blood pressure, and other lifestyle-associated risk factors in diabetes, using green (low risk), yellow (moderate risk), and red (high risk) indicators during diabetes consultations. The level of risk is then discussed with the patient in relation to actions that they can take to reduce their risk and activities they can undertake as part of self-management based on personal goals.

> *Here is a tool to facilitate the conversation with the patient, and this is important, and this is also prioritized. I think that's good.* [Participant, cycle 4]

The HCPs believe that the risk assessment tool can be used to engage persons with T2D and enhance their understanding of their health situations. Visualizing risk can increase motivation for self-management, thereby reducing cardiovascular risk. By asking open-ended questions about their own cardiovascular risk, persons with T2D can be encouraged to reflect on their situation, consider what actions they can take to reduce their risk, and bring attention to the emotional aspects of risk. For example: "What thoughts or feelings arise when you reflect on your risk?"

> *Is there anything you can do yourself to change this red number to yellow? It should be moving toward green, right?* [Participant, cycle 4]

The HCPs felt that the risk assessment tool was valuable for creating a person-centered plan that could prevent complications associated with T2D. The HCPs also argued that a color scale with numerical values would be useful for assessing risk.

> *Using colors to illustrate and stimulate thought processes is a great idea. It's important to have tools that can be prioritized in the conversation.* [Participant, cycle 4]

Some HCPs believed that the severity conveyed by such visuals might be important for encouraging persons with T2D to take risks seriously. However, some HCPs believed that the severity conveyed by such visuals might be unnecessarily worrying for persons with T2D and cause them to take the risks too seriously. These varying perspectives highlight the delicate balance between providing clear, impactful communication and avoiding unnecessary fear or discomfort.

> *I know that I may not always do everything the right way, because I can sometimes tend to want to comfort persons when they receive a diagnosis and say "we have such good medicines now" [laughs] Like, this will go well, it's not like before when people had to undergo amputations and suffered myocardial infarctions. So, one can risk sugarcoating things too much. But … for some, I believe that the … pale finger of death … might, in some way, help to bring about change [laughs] I believe that.* [Participant, cycle 4]

However, other findings revealed that some HCPs were concerned about using red to indicate a high level of risk, as they felt that it could be perceived as overly alarming or frightening:

> *And if they are curious and want to see this, it's great to show them, but not unthinkingly or at any time – especially not when they are in the red zone, because that could overwhelm them.* [Participant, cycle 4]

Material to Prompt Reflection

Material to prompt reflection was developed by the researchers for the diabetes specialist nurses to use in risk communication. The material consisted of questions intended to enhance understanding of the relationship between T2D and cardiovascular risk, encourage personal reflection on risk, create a safe space for discussing emotional aspects such as fear, uncertainty, and hope for the future, and promote motivation for self-management.

During previous cycles, it became apparent that risk communication might be insufficient. This finding deepened the researchers' understanding that diabetes specialist nurses may need additional support in their risk communication efforts. Consequently, the HCPs suggested that a risk-communication tool would be valuable in encouraging persons with T2D to share their thoughts. They also considered the reflection material

XSL•FO
RenderX

to be important for communicating regarding cardiovascular risk. The participants suggested that the reflection material should be discussed during consultations to capture emotional responses to risk:

> *It sounds wiser that they receive it together with the nurse so that they can ask questions directly, instead of taking home red values and then losing all motivation, thinking that the visit wasn't enjoyable at all, and deciding to just ignore it all.* [Participant, cycle 4]

Web Education for Diabetes Specialist Nurses

The results from previous cycles revealed that the diabetes specialist nurses need to recognize the importance of adopting a more person-centered approach, actively exploring and engaging with patients' own perceptions of risk. This includes asking questions that prompt reflection on how patients perceive their personal risk and emotional aspects, such as how they feel and whether they experience anxiety when discussing potential risks. To support diabetes specialist nurses in risk communication, web education was developed.

The web education aims to improve person-centered risk communication. The content includes the description and use of supportive tools, such as a patient handbook and a risk assessment tool developed by the researchers through a co-creation process. The web education also encourages critical reflection on the role of HCPs in risk communication and provides links to relevant guidelines and scientific literature. It also incorporates visual illustrations that show how to communicate risk in a person-centered way during consultations.

During cycle 4, the HCPs felt that the web education was highly supportive for diabetes specialist nurses, meeting their needs because it can be used in a time-efficient manner. It was perceived as flexible, allowing the diabetes specialist nurses to complete each module at their own pace and suit their schedules, making it easier to integrate the training with work commitments. As one physician explained:

> *I think that's fantastic, to get diabetes nurses to not feel that it's burdensome, but to feel that they can get help and be inspired. I believe that can make an incredible difference and help them not feel alone.* [Participant, cycle four]

Discussion

Principal Findings

The co-creation research design was intended to give the participants a strong voice, with the researchers facilitating the participants during the workshops. This aligns with previous research showing that participants actively shape the research, rather than merely contributing insights, and that creativity and innovative thinking are central to co-creation [24]. Co-creation also strengthens trust and relationships between researchers and participants [24]. This trust is not only a prerequisite for successful collaboration but an outcome of the process. When participants feel heard and respected, relationships between researchers and participants are strengthened, and different perspectives are presented. Hence, new ideas and solutions often emerge that may not have been possible in a more hierarchical research setup.

This co-creation study involved persons with T2D and HCPs. One review has highlighted that co-creation is often used to enhance health outcomes [25]. Our co-creation process revealed insufficient person-centered risk communication. The workshops provided insights into needs, barriers, and co-creative solutions and content for interventions, such as a personal handbook, a risk assessment tool, reflection materials, and web-based education for diabetes specialist nurses. These risk communication tools are designed to drive behavior change in line with self-determination theory [26]. This co-creation intervention supports person-centered communication and self-management to reduce cardiovascular risk in T2D and has the potential to increase risk awareness in persons with T2D.

In this study, we differentiate between PAR and design thinking. Although both approaches emphasize participation, empathy, and practice-oriented knowledge development, their underlying views of knowledge, understandings of power, and goals for change differ [27]. PAR can be seen as a transformative knowledge practice because it changes relationships and structures, whereas design thinking is an innovation-focused design practice that aims to improve solutions within existing frameworks [27]. The key difference between the two lies not in the level of participation but in the purpose of participation and how power is understood and distributed during the process. Adopting a design thinking approach might have risked reducing the participants' roles to those of users or test participants within a predefined problem framework.

The CFIR guided the development of an intervention due to constructs such as adaptability, patient needs and resources, and knowledge and beliefs about the innovation [28]. The insights helped develop strategies and materials to improve risk communication. In the "redefine and retest" phase, these were turned into a patient handbook, web-based educational resources, a risk assessment tool, and revised reflection materials. Overall, co-creation consistently guided the intervention to align with CFIR constructs (Multimedia Appendix 2).

Taking the person's voice into account showed that language barriers and overly medicalized terminology may hinder the understanding of cardiovascular risk in T2D. However, some research suggests that HCPs prefer using medical risk tools in diabetes communication [29]. This highlights the need to balance factual communication and empathetic understanding. HCPs must recognize that both cognitive and emotional aspects influence the willingness of patients to engage with risk communication. Some persons with T2D are unaware that diabetes can lead to myocardial infarction or stroke, highlighting the need for improved risk awareness. Additionally, those living with both T2D and cardiovascular disease as a multimorbidity do not perceive the conditions to be connected and view them as 2 separate conditions [30]. This underscores the importance of developing risk communication to support persons with T2D in gaining a deeper understanding of the seriousness of their condition and improving awareness of the crucial role that self-management plays in mitigating cardiovascular risk in T2D.

XSL•FO
RenderX

It is crucial to communicate risk to patients in relation to their own perspective on their disease and perceived risk, as this approach is essential to enhancing person-centered risk communication and supporting effective self-management.

Problem-solving and idea generation during the workshops revealed valuable insights to support person-centered approaches in risk communication consultations. The findings highlighted the need for HCPs to actively explore the emotional aspects of T2D and to use individualized, verbal, and written communication that is more direct, less vague, and accessible, so that it is easier for persons with T2D to understand and refer to. Persons with T2D emphasized the importance of undivided attention, emotional presence, and meaningful dialogue. Research shows that stigma associated with T2D can negatively affect the communication efforts of HCPs by, for example, their attributing blame and pointing to personal shortcomings such as a lack of willpower, discipline, or motivation to improve one's health. This can lead to negative perceptions of persons with T2D, as well as reduced confidence in their ability to communicate in a person-centered manner, and less effective motivational conversations [31]. A person-centered approach has been shown to reduce diabetes-related stigma, which in turn can facilitate improved risk communication [32].

This study demonstrated that co-creation can be used to develop a wide range of prototypes and interventions. The workshops gave persons with T2D and HCPs an active role in shaping an intervention, rather than being passive recipients of it. The knowledge of the participants regarding communication of cardiovascular risk was improved throughout the workshops, aligning with previous research that suggested that knowledge fosters engagement and creativity [33]. The co-creation process emphasized shared understanding and learning, with participant feedback being crucial to the development of the intervention. The focus was on developing the intervention *with*, rather than *for*, persons with T2D and HCPs. Additionally, the findings of this research support the conclusions of other studies: involving end users in intervention development increases the likelihood of successful implementation of the intervention in clinical practice [16].

In the redefine and retest stages, the participants reviewed the proposed intervention and suggested improvements. Some HCPs were hesitant to use emotionally charged colors such as red, fearing that these might alarm persons with T2D or be perceived as aggressive. In contrast, patients expressed a desire for clearer, more direct risk communication; for example, in workshops 1 and 2, persons with T2D expressed that they want to know their personal CVD risk, which is also illustrated in the study of Jutterström [17].

Although a cautious approach to communication is important, excessive caution may hinder effective risk communication, thereby affecting patient understanding and engagement. The results also indicate a need for deeper exploration of emotions during consultations, an aspect that is often overlooked in traditional health care [34]. The developed intervention may support diabetes specialist nurses in communicating the emotional and existential aspects of T2D and cardiovascular risk more clearly [25]. Greater attention to emotional dimensions may help integrate medical knowledge with emotional well-being, creating opportunities for more person-centered communication [34]. To address this, we developed a web-based educational program for diabetes specialist nurses that promotes person-centered risk communication, encouraging patients to articulate their personal perceptions of risk. Furthermore, the risk assessment document will be tested in a pilot study. Health care professionals will be informed that they may refrain from using the instrument if they judge it to be inappropriate for a particular patient.

Future research will test and evaluate the feasibility of the intervention in a pilot study conducted in primary health care. This aligns with findings that emphasize the need for health care interventions to fit the specific structures and resources of the health care systems within which they are to be utilized [9]. Integrating the intervention into existing care pathways could enhance accessibility and sustainability, making it a practical tool for supporting risk communication.

Study Limitations

Conducting separate workshops for patients and health care professionals prevented direct dialogue and limited insights into power dynamics and interactions. The results of previous workshops were based on earlier discussions that enabled participants to share their ideas. Additionally, our aim was to prevent a potential power imbalance [21]. It is uncertain whether the patients' views would have been fully represented in combined workshops.

Despite the small number of HCPs involved, the workshops enabled reflective and personal discussions, fostering trust and producing rich data. Small groups allowed the participants to express and develop their ideas freely. Reflexive thematic analysis, which does not require data saturation, supports the value of such rich, qualitative input, even from limited samples [23]. However, the small and potentially homogeneous group could limit transferability [35].

A key strength was the iterative co-creation process, in which participant feedback from earlier workshops was continuously used to refine the intervention, ensuring that real-world needs were met. A potential bias may have occurred as some participants were particularly familiar or comfortable with the topic, which could have influenced the depth or direction of discussions. Recruitment challenges led to a shift from purposeful to convenience sampling, possibly reducing sample diversity. This convenience-based method may have introduced selection bias, potentially favoring participants who were already more engaged in their care. However, this can be seen as a strength, as it aims to develop and improve diabetes care.

A limitation is that the developed person-centered intervention may be more suitable for patients who can engage with written information and reflective discussions. It may be less suitable for those with language or cognitive barriers, or with limited capacity to engage with written and reflective material.

In-person workshops with persons with T2D supported empathetic, in-depth dialogue, while digital workshops suited time-constrained HCPs and enabled broader participation. Separate groups minimized power imbalances and encouraged

XSL•FO
RenderX

openness, especially among persons with T2D. However, the different formats made comparisons difficult and probably prevented direct dialogue, which could have fostered mutual understanding and integration. When comparing the 4 workshops, the face-to-face sessions encouraged more spontaneous dialogue and the opportunity to meet other patients with T2D. In contrast, the digitally facilitated sessions were more practical and made participation easier for HCPs. Despite these differences, the feedback revealed similar discussions, suggesting that the results are comparable across different formats.

Through the PAR framework, the researchers positioned the persons with T2D and HCPs as active co-creators. Together, these participants developed an intervention that was intended to increase risk awareness among persons with T2D and provide them with greater control over their health. This may eventually lead to more sustainable and positive outcomes in their disease management. However, while PAR enhances relevance and collaboration, it is time- and resource-intensive, posing challenges related to tight timelines and funding constraints [13].

Conclusions

Using a PAR framework, this study highlighted the value of co-creation in developing an intervention aiming to enhance cardiovascular risk awareness among persons with T2D. The results showed that diabetes specialist nurses need to be encouraged to explore patients' perceptions of risk and create space to discuss emotional responses to risk during consultations. The co-created intervention consists of several components: a patient handbook, which supports patients in exploring their personal understanding of cardiovascular risk; a risk assessment tool, which visualizes individual risk factors based on laboratory values and facilitates discussions regarding risk during diabetes consultations; materials to prompt reflections to explore personal understanding of cardiovascular risk; and a web education for diabetes specialist nurses, which is designed to strengthen the person-centered approach. Web education is intended to encourage active exploration of patients' perceptions of risk, including the emotional and existential dimensions of risk, and to guide the use of various risk-communication tools during patient encounters. The developed intervention provides a practical example of how HCPs can implement person-centered communication about cardiovascular risk in persons with T2D. By integrating the intervention in diabetes care, a person-centered approach could be facilitated, and individuals will be supported in identifying intrinsic motivation for lifestyle changes. This co-created intervention has the potential to increase risk awareness among persons with T2D.

Acknowledgments

We want to thank all the persons with type 2 diabetes, health care professionals, physicians, and diabetes specialist nurses who participated in this study.

The language model Microsoft 365 Copilot [36] (provided by Umeå University) was used as a language-editing service during the writing process of the study to check grammar and spelling. Artificial intelligence was not used to perform core research tasks, such as generating scientific insights, analyzing and interpreting data, and drawing conclusions.

Funding

This study was funded by a regional agreement between Umeå University and Region Västerbotten (ALF), the Healthcare Region of Northern Sweden (VISARE NORR), the Swedish Diabetes Foundation, and the Medical Foundation in Skellefteå.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of the analysis process.
[DOCX File, 30 KB - diabetes_v11i1e85748_app1.docx]

Multimedia Appendix 2
The Consolidated Framework for Implementation Research (CFIR) and its relation to the study results.
[DOCX File, 29 KB - diabetes_v11i1e85748_app2.docx]

Checklist 1
RTARG checklist.
[DOCX File, 28 KB - diabetes_v11i1e85748_app3.docx]

XSL•FO
RenderX

References

1. Yang R, Xu H, Pedersen NL, et al. A healthy lifestyle mitigates the risk of heart disease related to type 2 diabetes: a prospective nested case-control study in a nationwide Swedish twin cohort. Diabetologia 2021 Mar;64(3):530-539. [doi: 10.1007/s00125-020-05324-z] [Medline: 33169206]
2. Song DK, Hong YS, Sung YA, Lee H. Risk factor control and cardiovascular events in patients with type 2 diabetes mellitus. PLoS One 2024;19(2):e0299035. [doi: 10.1371/journal.pone.0299035] [Medline: 38422102]
3. ElSayed NA, Aleppo G, Aroda VR, et al. 10. Cardiovascular disease and risk management: standards of care in diabetes-2023. Diabetes Care 2023 Jan 1;46(Suppl 1):S158-S190. [doi: 10.2337/dc23-S010] [Medline: 36507632]
4. Schulberg SD, Ferry AV, Jin K, et al. Cardiovascular risk communication strategies in primary prevention. A systematic review with narrative synthesis. J Adv Nurs 2022 Oct;78(10):3116-3140. [doi: 10.1111/jan.15327] [Medline: 35719002]
5. Bakhit M, Fien S, Abukmail E, et al. Cardiovascular disease risk communication and prevention: a meta-analysis. Eur Heart J 2024 Mar 27;45(12):998-1013. [doi: 10.1093/eurheartj/ehae002] [Medline: 38243824]
6. Rutten GEHM, Van Vugt H, de Koning E. Person-centered diabetes care and patient activation in people with type 2 diabetes. BMJ Open Diabetes Res Care 2020 Dec;8(2):bmjdrc-bm2020. [doi: 10.1136/bmjdrc-2020-001926] [Medline: 33323460]
7. Vornanen M, Konttinen H, Peltonen M, Haukkala A. Diabetes and cardiovascular disease risk perception and risk indicators: a 5-year follow-up. Int J Behav Med 2021 Jun;28(3):337-348. [doi: 10.1007/s12529-020-09924-2] [Medline: 32808182]
8. Gidlow CJ, Ellis NJ, Cowap L, et al. Quantitative examination of video-recorded NHS Health Checks: comparison of the use of QRISK2 versus JBS3 cardiovascular risk calculators. BMJ Open 2020 Sep 25;10(9):e037790. [doi: 10.1136/bmjopen-2020-037790] [Medline: 32978197]
9. Mayo NL, Russell HA, Holt K, Williams GC. Implementation of a self-determination based clinical program to reduce cardiovascular disease risk. J Health Psychol 2022 Nov;27(13):2898-2908. [doi: 10.1177/13591053211068124] [Medline: 35086378]
10. Kuipers SJ, Nieboer AP, Cramm JM. Easier said than done: healthcare professionals' barriers to the provision of patient-centered primary care to patients with multimorbidity. Int J Environ Res Public Health 2021 Jun 4;18(11):6057. [doi: 10.3390/ijerph18116057] [Medline: 34199866]
11. Pizga A, Karatzanos E, Tsikrika S, et al. Psychosocial interventions to enhance treatment adherence to lifestyle changes in cardiovascular disease: a review of the literature 2011-2021. European J Env Publi 2022;6(1):em0102. [doi: 10.21601/ejeph/11582]
12. O'Cathain A, Croot L, Sworn K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41. [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
13. Vargas I, Mogollón-Pérez AS, Eguiguren P, et al. Lessons learnt from the process of designing care coordination interventions through participatory action research in public healthcare networks of six Latin American countries. Health Res Policy Syst 2023 Jun 1;21(1):39. [doi: 10.1186/s12961-023-00985-9] [Medline: 37264416]
14. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
15. Hall J, Morton S, Hall J, et al. A co-production approach guided by the Behaviour Change Wheel to develop an intervention for reducing sedentary behaviour after stroke. Pilot Feasibility Stud 2020;6:115. [doi: 10.1186/s40814-020-00667-1] [Medline: 32821421]
16. Cornish F, Breton N, Moreno-Tabarez U, et al. Participatory action research. Nat Rev Methods Primers 2023;3(1). [doi: 10.1038/s43586-023-00214-1]
17. Jutterström L, Stenlund AL, Otten J, Lilja M, Hellström Ängerud K. Awareness of cardiovascular risk among persons with type 2 diabetes: a qualitative study. Int J Qual Stud Health Well-being 2024 Dec;19(1):2294512. [doi: 10.1080/17482631.2023.2294512] [Medline: 38112175]
18. Hayvon JC. Participatory theory building: a methodological review and conceptualization. Am J Health Educ 2025 May 4;56(3):266-277. [doi: 10.1080/19325037.2025.2466963]
19. Brydon-Miller M, Kral M, Ortiz Aragón A. Participatory action research: international perspectives and practices. Int Rev Qual Res 2020 Aug;13(2):103-111. [doi: 10.1177/1940844720933225]
20. Pearce G, Magee P. Co-creation solutions and the three Co's framework for applying co-creation. Health Educ 2024 Jun 6;124(1-2):20-37. [doi: 10.1108/HE-09-2022-0077]
21. Hadjiconstantinou M, Dunkley AJ, Eborall H, Robertson N, Khunti K, Davies M. Perceptions of healthcare professionals and people with type 2 diabetes on emotional support: a qualitative study. BJGP Open 2020;4(1):bjgpopen20X101018. [doi: 10.3399/bjgpopen20X101018] [Medline: 32184215]
22. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
23. Braun V, Clarke V. Supporting best practice in reflexive thematic analysis reporting in Palliative Medicine: a review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). Palliat Med 2024 Jun;38(6):608-616. [doi: 10.1177/02692163241234800] [Medline: 38469804]

XSL•FO
RenderX

24. Agnello DM, Anand-Kumar V, An Q, et al. Co-creation methods for public health research—characteristics, benefits, and challenges: a Health CASCADE scoping review. BMC Med Res Methodol 2025 Mar 6;25(1):60. [doi: 10.1186/s12874-025-02514-4] [Medline: 40050729]
25. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 7;22(1):877. [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
26. Ntoumanis N, Ng JYY, Prestwich A, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
27. Hahn-Goldberg S, Chaput A, Rosenberg-Yunger Z, et al. Tool development to improve medication information transfer to patients during transitions of care: a participatory action research and design thinking methodology approach. Res Social Adm Pharm 2022 Jan;18(1):2170-2177. [doi: 10.1016/j.sapharm.2021.04.002] [Medline: 33926826]
28. Reardon CM, Damschroder LJ, Ashcraft LE, et al. The Consolidated Framework for Implementation Research (CFIR) user guide: a five-step guide for conducting implementation research using the framework. Implement Sci 2025 Aug 16;20(1):39. [doi: 10.1186/s13012-025-01450-7] [Medline: 40819067]
29. Rouyard T, Leal J, Salvi D, Baskerville R, Velardo C, Gray A. An intuitive risk communication tool to enhance patient-provider partnership in diabetes consultation. J Diabetes Sci Technol 2022 Jul;16(4):988-994. [doi: 10.1177/1932296821995800] [Medline: 33655766]
30. Sebastian SA. Cardiovascular disease risk communication: strategies, impact, and future directions. Curr Probl Cardiol 2024 May;49(5):102490. [doi: 10.1016/j.cpcardiol.2024.102490] [Medline: 38417473]
31. Brown JE, Newson L, Umeh K, Lotto R, Jones I. Living with a multimorbidity. A qualitative study on the personal perspectives of individuals with type 2 diabetes and cardiovascular disease. Psychol Health 2025 Aug 19;2025:1-27. [doi: 10.1080/08870446.2025.2546411] [Medline: 40827896]
32. Bennett BL, Puhl RM. Physicians' stigmatizing attitudes about individuals with type 2 diabetes: associations with communication practices and perceived barriers to care. Prim Care Diabetes 2024 Oct;18(5):518-524. [doi: 10.1016/j.pcd.2024.07.001] [Medline: 39048399]
33. Leask CF, Sandlund M, Skelton DA, et al. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. Res Involv Engagem 2019;5:2. [doi: 10.1186/s40900-018-0136-9] [Medline: 30652027]
34. Muntinga M, Bendien E, Abma T, Groot B. Participatory action research and intersectionality: a critical dialogical reflection of a study with older adults. Qual Ageing Older Adults 2024 Feb 20;25(1):9-20. [doi: 10.1108/QAOA-03-2023-0024]
35. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
36. Microsoft 365 Copilot [large language model]. Microsoft. URL: https://www.microsoft.com/sv-se/microsoft-365-copilot [accessed 2026-04-04]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
HbA_{1c}: hemoglobin A_{1c}
HCP: health care professional
PAR: participatory action research
T2D: type 2 diabetes

Edited by R Zhao; submitted 12.Oct.2025; peer-reviewed by D Moreno-Martinez, K McCauley; revised version received 22.Feb.2026; accepted 15.Mar.2026; published 23.Apr.2026.

Please cite as:
Stenlund AL, Hellström Ängerud K, Lilja M, Otten J, Jutterström L
The Process of Developing an Intervention to Increase Awareness of Cardiovascular Risk for Persons With Type 2 Diabetes: Co-Creation Study
JMIR Diabetes 2026;11:e85748
URL: https://diabetes.jmir.org/2026/1/e85748
doi:10.2196/85748

© Anna-Lena Stenlund, Karin Hellström Ängerud, Mikael Lilja, Julia Otten, Lena Jutterström. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 23.Apr.2026. This is an open-access article distributed under the terms of the Creative

XSL•FO
RenderX

Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personalized Digital Health Solutions to Increasing Diabetes-Related Knowledge and Behavioral Outcomes: Results From a Randomized Controlled Trial

Marcia Ory[1,2*], MPH, PhD; Christi H Esquivel[1*], MEd, PhD; Gang Han[1,3], PhD; Samuel D Castiglione Towne Jr[1,2,4,5], MPH, PhD; SangNam Ahn[1,6], MPSA, PhD; Matthew Lee Smith[1,7,8,9], MPH, PhD

[1]Center for Community Health and Aging, School of Public Health, Texas A&M University, 212 Adriance Lab Rd, College Station, TX, United States
[2]Department of Environmental and Occupational Health, School of Public Health, Texas A&M University, College Station, TX, United States
[3]Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M University, College Station, TX, United States
[4]Disability, Aging, and Technology Cluster, University of Central Florida, Orlando, FL, United States
[5]School of Global Health Management and Informatics, University of Central Florida, Orlando, FL, United States
[6]Department of Health Management and Policy, College for Public Health and Social Justice, Saint Louis University, Saint Louis, MO, United States
[7]Department of Health Behavior, School of Public Health, Texas A&M University, College Station, TX, United States
[8]Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States
[9]Center for Health by Design, School of Public Health, Indiana University, Bloomington, IN, United States
[*]these authors contributed equally

Corresponding Author:
Christi H Esquivel, MEd, PhD
Center for Community Health and Aging, School of Public Health, Texas A&M University, 212 Adriance Lab Rd, College Station, TX, United States

Abstract

Background: The prevalence of diabetes in the United States necessitates investigations into how to better enable adults with type 2 diabetes mellitus (T2DM) to manage their health using easy-to-access and personally adaptable technologies. The ubiquity of digital content further justifies the need to consider the impact of different digital intervention modalities in diabetes self-care activities.

Objective: This study aimed to compare the impact of 2 digital diabetes self-care education programs delivered separately and in combination to adults with T2DM across various settings in Texas.

Methods: We conducted a randomized controlled trial in Texas with 188 adults with T2DM to assess whether 2 different interventions alone (Virtual Making Moves with Diabetes or Technology-Based Education and Support) or in combination (Virtual Making Moves with Diabetes followed by Technology-Based Education and Support) improved multiple outcomes associated with diabetes self-management. We used several estimation techniques, including generalized estimating equations, to account for multiple factors simultaneously.

Results: All 3 digital intervention modalities led to statistically significant improvements in diabetes-related confidence, distress, and self-care behaviors, with significance from baseline through 6 months and supported by moderate to strong effect sizes (Cohen *d*) ranging from 0.446 to 0.827 at 3-month follow-up versus baseline and from 0.538 to 0.888 at 6-month follow-up versus baseline. No statistically significant superiority was observed among the intervention modalities. Higher self-care behaviors were significantly associated with higher baseline confidence and lower distress. Those in the most disadvantaged positions (less education, less financial stability, and no health insurance) showed significantly larger improvement in self-care behaviors.

Conclusions: Given the benefits associated with this study's interventions, we suggest future work to further develop digital content that can be tailored to individuals with T2DM to help them manage their chronic conditions in a cost-effective manner.

Trial Registration: ClinicalTrials.gov NCT06370494; https://clinicaltrials.gov/study/NCT06370494

(JMIR Diabetes 2026;11:e87364) doi:10.2196/87364

KEYWORDS

diabetes self-management education and support; diabetes self-care behaviors; digital health; randomized clinical trial; diabetes knowledge; diabetes confidence; diabetes distress

XSL•FO
RenderX

Introduction

Background

Type 2 diabetes mellitus (T2DM) is complex, costly, and often preventable; yet, millions still live with this destructive disease, with many more Americans diagnosed annually. Ideally, incidence rates would be on a downward trend; however, this is not the case, with adult incidence rates remaining relatively stable over the past ~20 years, albeit with some variation following 2008 [1]. Thus, with more individuals diagnosed, coupled with this being a lifelong disease upon diagnosis, prevalence rates are on the rise, with an increase from 10% (2001 - 2004) to 13% (2017 - 2020) [1]. Therefore, this remains a critical national public health issue with additional evidence of differential patterns impacting individuals across both individual-level and place-based characteristics [2].

The United States is home to many residents living outside major metropolitan areas, who have historically faced inequities in accessing critical health care resources [3,4]. Recent estimates suggest that age-adjusted prevalence of diagnosed diabetes was higher among those residing in nonmetropolitan areas, with rates of 8.9% and 7.9% for both men and women, respectively, among large central metro areas, compared to rates of 10.5% and 8.6% in nonmetropolitan areas [1]. Furthermore, in terms of individual-level characteristics, inequities exist across characteristics such as age (with higher rates with increasing age) and sex (with higher rates among males) [1]. Thus, this is a complex and interrelated issue spanning multiple levels. Making matters worse, individuals with diabetes often face potentially preventable medical encounters due to, in part, poor management of diabetes, including higher risks of heart disease, stroke, blindness, kidney disease, amputations, and other items [5]. For many, these complications could potentially be prevented with proper maintenance of diabetes and related comorbidities, as applicable.

Effective self-management involves taking the necessary steps to monitor and maintain optimal glycemic levels and prevent adverse effects [6-8]. Additional components of effective self-management include medication adherence and lifestyle modifications (eg, diet or nutrition modifications, physical activity, and refraining from, or quitting, smoking) [9-11]. Thus, interventions aimed at reducing potentially preventable complications and improving the quality of life of individuals diagnosed with diabetes should consider both individual-level characteristics in combination with place-based measures, consistent with a strong theoretical foundation [12,13].

In terms of interventions, they can take several forms and platforms, including but not limited to online access to critical information; yet, there are some core elements to consider. For example, experts recommend all patients diagnosed with diabetes receive diabetes self-management education and support (DSMES) to provide them with the knowledge, resources, and skills essential to support their self-care journey and manage their condition [8]. Research indicates patients who receive DSMES are more likely to practice recommended self-care behaviors and clinical preventive care practices, such as routine clinical examinations and vaccines [9]. Additionally, positive changes in knowledge [14,15], confidence [8,16], and reduced distress [8,17] have been consistently reported and associated with sociodemographic and health factors, such as age, sex, education, race or ethnicity, rurality, financial stability, insurance coverage, and general health status [18-20].

With the rise in technological development and aftereffects of the COVID-19 pandemic, there is now a greater variety and availability of DSMES in the virtual and digital space. Interventions come in various formats from in-person to virtual (including both synchronous and asynchronous) to hybrid and contain various components or mechanisms of support, including, but not limited to, behavioral tracking and monitoring, peer support, coaching, knowledge-based resources, and skill building [21-23]. As no program is a one-size-fits-all, there are pros and cons to consider for both in-person and virtual DSMES programs.

Consistent with the diffusion of innovations theory [24], many people may be skeptical about the effectiveness of new digital health interventions (which include telehealth, telemedicine, mobile health, and eHealth [25]) and wary to engage in such practices. However, a systematic literature review and meta-analysis revealed that digital health interventions for people with T2DM were associated with significant improvements in hemoglobin A_{1c} levels [26]. Additionally, digital interventions may fill some gaps of in-person interventions, including accessibility, real-time feedback, and transportation barriers [27,28]. However, it is also important to note that users' health literacy and/or digital literacy may impact their usage, level of engagement, and outcomes of using digital health interventions [29,30].

As new programs are developed, researchers have analyzed the acceptability, usability, and impacts of digital DSMES [7]. However, most studies analyzing the impacts and outcomes of DSMES compare the use of a digital program to a control group [7,31-35]. Additionally, studies tend to recruit from one community, limiting the generalizability of findings and the inability to observe the effectiveness of different digital health DSMES programs on behavioral outcomes across different geographic regions.

Study Aims

This study, part of the Texas A&M University Rural Health Moonshot Initiative [6], attempted to address these research gaps by comparing the impact of 2 digital DSMES programs delivered separately and in combination to adults with T2DM across many settings in Texas. The objectives were to examine (1) the sociodemographic and general health characteristics of adults living with diabetes who agreed to participate in a randomized clinical trial and whether any of these baseline characteristics were significantly associated with group assignment which served as a test of our randomization protocol; (2) mean changes and effect size differences in knowledge, confidence, distress, and self-care activities by intervention and rurality status; and (3) whether knowledge, confidence, and distress, controlling for sociodemographic factors and general health status predicted changes in self-care activities over time.

XSL•FO
RenderX

Methods

This study is reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) reporting guidelines for randomized trials (Checklist 1).

Setting and Population

As described in detail in an earlier study [6], we recruited 188 adults aged 25 years and older living in Texas from November 2020 to March 2022 who reported having T2DM with baseline hemoglobin A_{1c} (HbA_{1c}) levels of 7.5 or higher, indicating poorly managed diabetes and potential benefit from an intervention [36,37] targeted toward improving one's management of diabetes. Recruitment efforts were geared toward participants from both rural and urban communities in Texas. Interested participants were considered eligible if they had access to a smartphone with internet connectivity and were able to read and speak English.

Patient and Public Partnership or Involvement

This trial was generated based on the research team's experiences of implementing diabetes management programming in rural and underserved communities [6,38,39]. Feedback was obtained from program implementers and participants in relation to program content, access, and attendance. Using this information obtained in the years preceding the trial, we updated the intervention content and delivery format accordingly. While originally attempting to test the in-person version of the intervention, the COVID-19 pandemic required us to pivot to virtual delivery. Representatives from community-based organizations, health care organizations, program facilitators, and past participants were engaged in the material development and testing phase of the digital translation. Furthermore, these representatives were engaged when attempting to identify the best recruitment and retention strategies, given high levels of attrition in past initiatives. Community and clinical partners were instrumental in participant recruitment efforts. While participants were not directly involved in the selection of measures, previously used subjective measures were included based on their success in related interventions and trials [6,38]. Findings were shared with the funder and community partners in accordance with their preferences. These took the form of informal meetings and discussions with stakeholders, flyers, virtual correspondence (eg, websites and social media), conference presentations, and publications.

Recruitment Strategies

Knowing that multiple strategies would be needed to reach community-dwelling participants throughout the state, a variety of recruitment strategies were used, including radio advertisements, radio segments, flyers, bulletin boards, social media campaigns, directed letter campaigns from in-network health care providers, newspapers, email listserves, in-person recruitment, and the study's website. We also used multiple retention strategies: participant incentives, distribution of study information, and frequent communication via text messaging, emails, and phone calls to limit attrition.

Ethical Considerations

This study was reviewed and approved by the Texas A&M University Institutional Review Board (IRB 2019-0804D) and registered at ClinicalTrials.gov (ID NCT06370494). All participants provided written consent to participate in this randomized controlled trial (RCT). Participation was voluntary, and participants could withdraw from the study at any time without penalty. Deidentified data on major study outcomes were placed in the Texas A&M data repository upon completion of data analyses and report writing. Only the study team could access study data. Depending on their participation in the study, each participant could receive upward of US $375 for completing data throughout the study period.

Interventions

Participants were randomized into 1 of 3 intervention arms, as described in the following paragraph. Each intervention aligned with American Diabetes Association national standards for diabetes self-management and best practices for behavioral change [40]. DSMES is a continuous process designed to help individuals with diabetes gain the knowledge, skills, and confidence needed to manage their condition effectively [41]. Each DSMES may vary in components, objectives, and mechanisms, but there is consensus that goals of DSMES are to promote informed decision-making, encourage healthy self-care habits, enhance problem-solving abilities, foster collaboration with health care providers, and ultimately improve health outcomes, overall well-being, and quality of life [41,42]. Furthermore, the American Diabetes Association and International Diabetes Federation established guidelines for DSMES programs to not only be reimbursable but also recognized and accredited as evidence-based and effective programs.

Differing in modality, the 3 intervention arms (further described in Table 1) included (1) Virtual Making Moves with Diabetes (vMMWD) Group, (2) Technology-Based Education and Support (TBES) Group, and (3) a combined approach of vMMWD and TBES. Participants from all groups received an HbA_{1c} monitoring kit, a glucose meter, and supplies to test their blood sugar at home. On the basis of the prior literature documenting positive effects of DSMES [8,9,16,17,21-23,43-46], we hypothesized that all intervention strategies would have a beneficial effect on diabetes-related study outcomes, but that effects would be stronger for vMMWD over TBES, given the greater personalization, and that the combined intervention strategy (representing the strengths of each intervention modality) would have the strongest effects. We also hypothesized that participants from traditionally underserved and resource-limited settings (eg, minority identity, less education, lower financial security, no health insurance, and rural residence) who typically have less access to diabetes self-care programs might have the greatest improvements.

XSL•FO
RenderX

Table . Overview of interventions.

Intervention group	Description
vMMWD[a] group	Asynchronous virtual education and support program with one-on-one follow-up counseling with a diabetes specialist. Entailed 3 education sessions within 3 mo (totaling about 6 - 8 h). Also entailed one-on-one interactions with a registered nurse or registered dietitian to personalize content, build upon virtual education sessions, and create tailored diabetes self-care strategies.
TBES[b] group	Smartphone app to provide continuous, but less structured, support. Participants used the app at their own pace to learn and practice diabetes self-care skills and strategies. The TBES included a chat feature for participants to get personalized support from a diabetes coach.
Combination group	Combined modality where participants sequentially received vMMWD and TBES (upon completion of vMMWD) to maximize support and behavioral reinforcement.

[a]vMMWD: Virtual Making Moves with Diabetes.

[b]TBES: Technology-Based Education and Support.

Measures

We selected standardized measures that have been previously used in diabetes education studies, as described below.

1. Diabetes knowledge: We used a variant of the Spoken Knowledge in Low Literacy in Diabetes Scale used in prior diabetes self-management studies [47,48]. Our knowledge measure was composed of 10 true or false knowledge questions, such as "Glycosylated hemoglobin (HbA_{1c}) is a test that measures your average blood sugar level in the past week" and "Having regular check-ups with your healthcare provider can help spot the early signs of diabetes complications." The total score represented the percentage of correctly answered questions.
2. Diabetes confidence: This was measured using the 8-item Self-Efficacy for Diabetes scale, rated on a 10-point scale from 1 ("not at all confident") to 10 ("totally confident"), with values ranging from 8 to 80 points, with higher points demonstrating better confidence in diabetes self-care management [49]. The diabetes self-efficacy scale measured participants' confidence to engage in self-care behaviors, such as eating meals every 4 to 5 hours, exercising for 15 to 30 minutes, 4 to 5 times a week, following their diet when they have to prepare or share food with other people who do not have diabetes, and knowing what to do when their blood glucose level goes higher or lower than it should be. An example item was as follows: "How confident do you feel that you can control your diabetes so that it does not interfere with the things you want to do?"
3. Diabetes distress: The Diabetes Distress Scale [50] was used to measure distress in 4 domains: emotional burden, physician-related distress, regimen-related distress, and interpersonal distress (eg, "Feeling that I am not testing my blood sugars frequently enough"). Six diabetes-specific items were rated on a 6-point Likert scale, ranging from 1 ("not a problem") to 6 ("serious problem"), with the scale ranging from 6 to 36 points. Reverse coding was not required, and a higher mean score of 6 items indicates greater levels of diabetes-related distress.
4. Diabetes self-care behaviors: We used 9 items from the Summary of Diabetes Self-Care Activities measure [51] based on the mean number of days engaged in different behaviors, such as general diet, specific diet, exercise, blood glucose testing, and foot care. The scale ranged from 0 to 63, with higher scores reflecting better engagement in key self-care behaviors. This scale has been shown to be a brief yet reliable and valid self-report measure of diabetes self-care used in both research and practice.
5. Sociodemographic and health-related variables: We included standard self-reported demographic measures of age, sex, race (White vs non-White), ethnicity (Hispanic or Latino vs not), education (more than high school vs high school or less), insurance coverage (yes vs no), and financial stability (yes vs no).
6. Place-based variables: Additionally, we characterized participants' settings as rural versus urban. The National Center for Health Statistics Urban–Rural Classification Scheme for Counties [52] was used to dichotomize participants into metropolitan (levels 1 - 4) and nonmetropolitan areas (levels 5 - 6) consistent with our aims of examining outcomes from both urban and rural areas.
7. Self-rated health: We also included a measure of self-rated health, asking participants to rate their health as excellent, very good, good, fair, or poor and grouped responses into 3 groups: excellent, very good or good, and fair or poor [53].

Study Design and Analyses

Overview

The team implemented an RCT to evaluate the relative effectiveness of three approaches: (1) vMMWD alone, (2) TBES alone, and (3) a combined sequence in which participants first received vMMWD followed by TBES. Each intervention was delivered over a 3-month period, with data collected at baseline and at 3- and 6-month follow-up. A 2-stage randomization process was used to allocate participants to the respective intervention arms, and no dropout was observed between the first and second randomization stages. Prior to study initiation,

XSL•FO
RenderX

sample size calculations were performed for the primary outcome measure, HbA_{1c}. Expected changes in HbA_{1c} were informed by findings from an earlier South Texas study using a similar intervention [38,39]. On the basis of an assumed moderate effect size (Cohen d=0.5), power analyses indicated that a minimum of 34 participants completing both baseline and postintervention assessments would be required to detect statistically significant differences in HbA_{1c}, with 80% power and a 2-sided paired t test at the 0.05 significance level. Assuming that effect sizes would be even larger for the self-care intervention impacts, the sample size was deemed adequate to detect statistically significant differences in self-care behaviors. As no nonintervention control group was included, all effectiveness analyses are interpreted as within-group changes over time. For between-group comparisons, the TBES-only group was treated as the reference category in regression models due to its lower intensity relative to other modalities.

Randomization Procedures

As described previously [6], the randomization involved 2 stages. The first stage was an individual-based block-stratified randomization into 1 of 2 arms, vMMWD only. or TBES only, with a ratio of 2:1, with 2 participants in vMMWD relative to 1 participant in TBES, which was stratified by rural and urban areas with a block size of 3. The second stage of randomization was conducted for the participants who completed vMMWD in the first stage. It was an individual-based block randomization for 1 of 2 arms: vMMWD only (ie, no additional intervention) versus combined modality (ie, receive TBES after they already received vMMWD), with a ratio of 1:1. Furthermore, in designing the study, we debated whether the combined treatment arm should be concurrent or sequential. For both logistical and conceptual reasons, we decided to phase in the TBES condition after the more personal vMMWD intervention, with the assumption that the app condition would help sustain any benefits achieved in a time-limited intervention.

All 3 intervention modalities were conducted for 3 months. Knowledge, confidence, distress, and self-care activity scores were observed at baseline and follow-up periods at 3 and 6 months.

Descriptive Analyses

As described in Ory et al [6] for the analysis of HbA_{1c}, an intent-to-treat analysis was also conducted for all the behavioral variables. Descriptive statistics were calculated for all participants by intervention arm. We provided medians with IQRs for continuous variables, and frequencies with percentages for discrete variables. Wilcoxon rank-sum tests and Pearson chi-square tests were used to test associations between low versus high baseline self-care and the covariates in Table 2 and regression.

Changes in knowledge, confidence, distress, and self-care activities were calculated as differences between baseline and 3-month follow-up and between baseline and 6-month follow-up (eg, “3 mo minus baseline” and “6 mo minus baseline”). For knowledge, confidence, distress, and self-care activity, the average changes and 95% CIs of mean changes based on t distributions were calculated for all participants, for participants in each study arm, and for participants residing in rural and urban areas. Using the average and SD of changes, we calculated and reported Cohen d from the paired t test as effect size.

Longitudinal Analyses

Longitudinal self-care activity scores at baseline and at 3-month and 6-month follow-ups were modeled as a longitudinal outcome using knowledge, confidence, and distress measured at baseline and follow-ups, as well as covariates from Table 2 measured only at baseline. The follow-up time (operationally defined as 0 at baseline, 3 at 3-month follow-up, and 6 at 6-month follow-up) was also a covariate in the analysis, which indicated whether the change of the outcome score was significant after adjusting for all the covariates. The fully adjusted model included follow-up time, knowledge, confidence, distress, treatment modality, and study demographic and health variables. The coefficients and P values of knowledge, confidence, and distress indicate the direction and significance of associations between these variables and the outcome variable. The longitudinal analysis with generalized estimating equations was implemented using SAS software (SAS Institute), the GENMOD procedure. Results from Tables 2–4 were generated using SAS software, version 9.4. A P value of .05 or less was considered statistically significant.

Causality

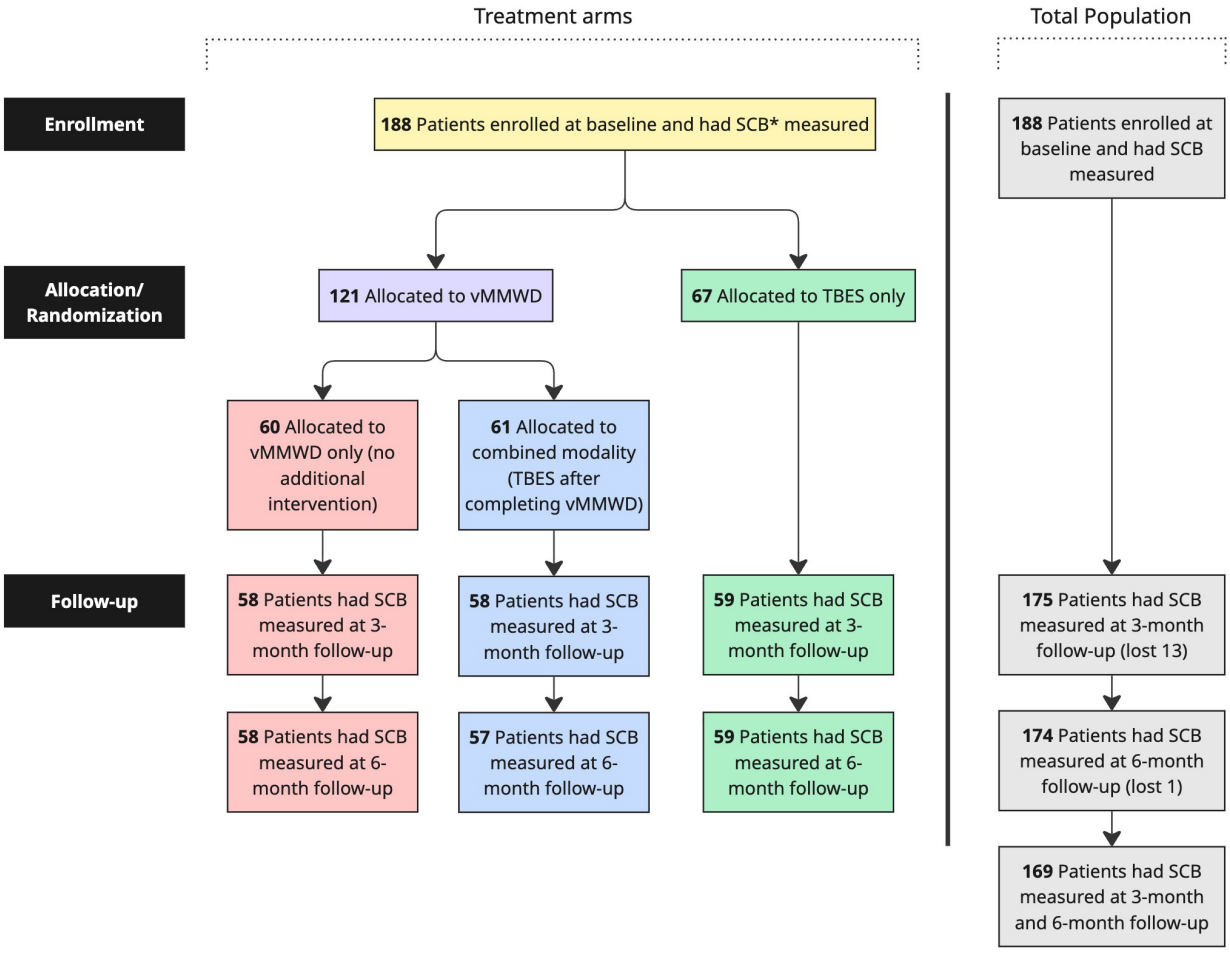
Random assignment was used to increase internal validity in terms of the intervention effects [54]. This increased our likelihood of identifying a true effect, considering the study design, albeit not without some limiting factors (see section *Limitations*).

Results

There was minimal participant attrition over time. As indicated in the CONSORT diagram (Figure 1), of the 188 participants enrolled at baseline, there was no drop off between enrollment and randomization. Concerns about selection bias based on the 2-stage randomization process were minimized because there were no significant differences between vMMWD completers and noncompleters prior to stage 2 randomization. Furthermore, the amount of missing data on the key dependent variable at 3-month and/or 6-month follow-ups was also minimal (ie, no more than 10%), mitigating the need for data imputation. Table 2 presents sample characteristics by intervention arm. In terms of the intervention arm, 32% (60/188) were in the vMMWD only condition, 36% (67/188) were in the TBES only condition, and 32% (61/188) were in the combined vMMWD and TBES condition. The median age of participants was 52 years (IQR 44-59), with most participants self-identifying as female (146/189, 77%), non-Hispanic or Latino (116/189, 61%), and White (163/189, 86%), and having more than a high school education (149/188, 79%). More than half of the participants (104/189, 55%) reported being in poor or fair health at baseline. At baseline, most participants reported being financially stable (160/189, 85%) and having health insurance (147/171, 86%). When comparing sample characteristics by intervention arm, no significant associations emerged (P>.05).

XSL•FO
RenderX

Figure 1. Participant flow diagram. SCB: self-care behavior; TBES: Technology-Based Education and Support; vMMWD: Virtual Making Moves with Diabetes.



XSL•FO
RenderX

Table . Sample characteristics by intervention arms.

Variables and level	All	vMMWD[a]	TBES[b]	Combined	*P* value
Age (y), median (range), n	52 (44-54), 189	50 (43-55), 60	53.5 (46.5 - 59.5) 68	53 (43-54), 61	.27
Sex, n (%)					
Female	146 (77)	47 (78)	54 (79)	45 (74)	.73
Male	43 (23)	13 (22)	14 (21)	16 (26)	
Ethnicity, n (%)					
Hispanic/Latino	73 (39)	25 (42)	23 (34)	25 (41)	.59
Not Hispanic/Latino	116 (61)	35 (58)	45 (66)	36 (59)	
Binary race, n (%)					
Non-White	26 (14)	11 (18)	7 (10)	8 (13)	.41
White	163 (86)	49 (82)	61 (90)	53 (87)	
General health, n (%)					
1=poor/fair	104 (55)	33 (55)	36 (53)	35 (57)	.17
2=very good/good	64 (34)	22 (37)	27 (40)	15 (25)	
3=excellent	21 (11)	5 (8)	5 (7)	11 (18)	
Rural status, n (%)					
Rural	67 (35)	22 (37)	23 (34)	22 (36)	.94
Urban	122 (65)	38 (63)	45 (66)	39 (64)	
Financial stability, n (%)					
0=no	29 (15)	10 (17)	10 (15)	9 (15)	.94
1=yes	160 (85)	50 (83)	58 (85)	52 (85)	
Education, n (%)					
0=less than high school	39 (21)	9 (15)	14 (21)	16 (26)	.31
1=greater than high school	149 (79)	51 (85)	53 (79)	45 (74)	
Have insurance, n (%)					
1=yes	147 (86)	50 (86)	50 (86)	47 (85)	.99
2=no	24 (14)	8 (14)	8 (14)	8 (15)	
Knowledge baseline, median (range), n	8 (7-9), 188	8 (7-9), 60	9 (7 - 9.5), 68	8 (6-9), 60	.23
Distress baseline, median (range), n	22 (14-28), 189	22 (14.5 - 29), 60	21 (14-27), 68	23 (15-27), 61	.79
Confidence baseline, median (range), n	45 (35-56), 189	41.5 (32.5 - 51.5), 60	49 (36 - 57.5), 68	45 (35-53), 61	.10
Self-care behavior baseline, median (range), n	26 (17-35), 188	21 (16.5 - 35.5), 60	27 (18-36), 67	27 (18-34), 61	.39

[a]vMMWD: Virtual Making Moves with Diabetes.

[b]TBES: Technology-Based Education and Support.

Table 3 reports mean changes in outcome variables overall and by rural status and intervention arm. Effect sizes are also reported, reinforcing that we were sufficiently powered to detect statistically significant differences. Overall, for all participants combined, significant improvements were observed from baseline to 3-month and 6-month follow-up for knowledge, confidence, distress, and self-care activities ($P<.001$). All but one subgroup (ie, the TBES only intervention arm) showed significant improvements from baseline to 3-month and 6-month follow-up for knowledge ($P<.001$). Among all geographic

XSL•FO
RenderX

statuses and intervention arms, significant improvements were observed from baseline to 3-month and 6-month follow-up for confidence, distress, and self-care activities ($P<.05$). Effect sizes for the total population ranged from 0.446 for knowledge to 0.827 for self-care activities at the 3-month follow-up and from 0.538 for knowledge to 0.888 for self-care activities at the 6-month follow-up assessment.

XSL•FO
RenderX

Table . Mean differences in outcome variables for all participants and by rural status and intervention arm.

Outcome variables	3 mo−baseline		6 mo−baseline	
	Mean Δ (95% CI) [n]	Effect size[a]	Mean Δ (95% CI) [n]	Effect size
Knowledge				
Total	0.769 (0.512 to 1.026)[b] [173]	0.446	0.866 (0.625 to 1.107)[b] [172]	0.538
Rural	0.921 (0.531 to 1.311)[b] [57]	0.583	0.922 (0.493 to 1.350)[b] [58]	0.527
Urban	0.682 (0.345 to 1.019)[b] [110]	0.378	0.833 (0.544 to 1.122)[b] [108]	0.544
TBES[c]	0.085 (−0.387 to 0.557) [54]	0.046	0.373 (−0.003 to 0.749) [54]	0.253
vMMWD[d]	1.086 (0.731 to 1.441)[b] [55]	0.787	1.121 (0.720 to 1.521)[b] [55]	0.720
Combined	1.161 (0.712 to 1.610)[b] [59]	0.677	1.127 (0.675 to 1.579)[b] [60]	0.659
Confidence				
Total	8.520 (6.211 to 10.830)[b] [171]	0.553	12.127 (9.724 to 14.530)[b] [173]	0.752
Rural	5.968 (2.051 to 9.884)[e] [61]	0.379	11.125 (6.907 to 15.343)[b] [58]	0.646
Urban	9.972 (7.138 to 12.807)[b] [109]	0.661	12.716 (9.806 to 15.626)[b] [109]	0.820
TBES	4.138 (0.029 to 8.247)[e] [55]	0.259	7.864 (4.203 to 11.526)[b] [54]	0.548
vMMWD	13.228 (9.418 to 17.038)[b] [62]	0.901	15.190 (11.418 to 18.961)[b] [55]	1.037
Combined	8.268 (4.501 to 12.035)[b] [59]	0.575	13.446 (8.600 to 18.293)[b] [59]	0.727
Distress				
Total	−4.392 (−5.477 to −3.307)[b] [171]	0.607	−5.509 (−6.913 to −4.106)[b] [106]	0.748
Rural	−4.145 (−5.863 to −2.428)[b] [61]	0.601	−4.500 (−7.011 to −1.989)[b] [30]	0.641
Urban	−4.532 (−5.931 to −3.133)[b] [109]	0.608	−5.908 (−7.597 to −4.219)[b] [63]	0.786
TBES	−2.517 (−4.321 to −0.714)[e] [55]	0.359	−3.568 (−6.020 to −1.115)[e] [37]	0.469
vMMWD	−6.404 (−8.363 to −4.444)[b] [62]	0.848	−7.353 (−9.767 to −4.939)[b] [34]	1.024
Combined	−4.286 (−6.046 to −2.526)[b] [59]	0.638	−5.771 (−8.086 to −3.456)[b] [35]	0.826
Self-care activities				
Total	9.994 (8.199 to 11.790)[b] [174]	0.827	10.295 (8.568 to 12.022)[b] [173]	0.888
Rural	9.143 (5.822 to 12.464)[b] [57]	0.680	11.125 (8.350 to 13.900)[b] [58]	0.982
Urban	10.477 (8.381 to 12.574)[b] [111]	0.930	9.807 (7.599 to 12.016)[b] [109]	0.834

XSL•FO
RenderX

Outcome variables	3 mo−baseline		6 mo−baseline	
TBES	8.431 (5.502 to 11.360)[b] [55]	0.741	7.638 (4.561 to 10.715)[b] [55]	0.639
vMMWD	12.759 (9.801 to 15.717)[b] [55]	1.110	12.948 (9.765 to 16.132)[b] [55]	1.047
Combined	8.793 (5.440 to 12.146)[b] [55]	0.675	10.298 (7.746 to 12.850)[b] [62]	1.048

[a]The effect size is the one-sample Cohen *d*.

[b]Statistical significance from testing if the change was different from 0 if the *P* value is less than .001.

[c]TBES: Technology-Based Education and Support.

[d]vMMWD: Virtual Making Moves with Diabetes.

[e]Statistical significance from testing if the change was different from 0 if the *P* value is between .001 and .05.

Table 4 reports findings from the longitudinal analysis examining factors associated with higher self-care activity scale scores at follow-up. The model was adjusted for the primary independent variable, follow-up time, and covariates, such as knowledge, confidence, distress, sociodemographic variables, and self-assessed health. On average, participants reported significantly higher self-care activity scores from baseline to follow-up (estimate=0.93; *P*<.001). Participants who reported higher baseline confidence scores (estimate=0.30; *P*<.001) and lower baseline distress scores (estimate=−0.50; *P*<.001) reported significantly higher self-care activity scores at follow-up, respectively. Participants who reported having less education (estimate=−3.10; *P*=.04), no health insurance (estimate=−3.35; *P*=.047), and lower financial stability at baseline (estimate=−3.93; *P*=.01) reported significantly higher self-care activity scores at follow-up, respectively. However, self-care activities were not significantly associated with diabetes knowledge or intervention modality (all *P*>.05).

XSL•FO
RenderX

Table . Longitudinal regression analysis of self-care activities with knowledge, confidence, distress, and covariates.

Parameters[a]	Estimate[b]	SE	95% CI		*P* value
			Upper limit	Lower limit	
Follow-up time					
From baseline to 6 mo	0.93	0.20	0.54	1.32	<.001
Knowledge					
Higher score reflects more correct answers	−0.37	0.39	−1.13	0.40	.35
Confidence					
Higher score reflects more confidence	0.30	0.05	0.20	0.39	<.001
Distress					
Higher score reflects more distress	−0.50	0.09	−0.68	−0.32	<.001
Age					
Continuously measured	0.02	0.07	−0.11	0.15	.76
Sex					
Female (vs male)	0.00	1.63	−3.18	3.19	.998
Ethnicity					
Hispanic (vs non-Hispanic)	0.05	1.49	−2.86	2.97	.97
Race					
White (vs non-White)	−0.74	1.92	−4.50	3.02	.699
Rural status					
Urban (vs rural)	−0.81	1.31	−3.37	1.75	.54
Education					
More than high school (vs high school or less)	−3.10	1.54	−6.11	−0.08	.04
General health					
Very good or good (vs fair or poor)	1.47	1.51	−1.49	4.43	.33
Excellent (vs fair or poor)	1.89	2.14	−2.29	6.08	.38
Arm					
Combined (vs app)	0.14	1.61	−3.01	3.29	.93
Class (vs app)	−0.83	1.44	−3.65	2.00	.57
Financial stability					
Ends meet (vs not met)	−3.93	1.61	−7.08	−0.78	.01
Insurance					
Yes (vs no)	−3.35	1.69	−6.66	−0.05	.047

[a]Referent levels are listed in parentheses.

[b]The positive estimate of time and confidence and negative estimate of distress are associated with improved self-care activities.

XSL•FO
RenderX

Discussion

Principal Findings

This study examined the comparative effectiveness of 3 digital DSMES interventions among adults with T2DM, with a particular focus on differences by intervention modality and sociodemographic factors. Findings demonstrated that (1) randomization was effective in creating 3 distinct intervention groups (objective 1); (2) all intervention modalities significantly improved diabetes-related confidence, distress, and self-care behaviors over time, with sustained effects (objective 2); and (3) confidence and distress, controlling for sociodemographic factors and general health status, predicted significant changes in self-care activities over time (objective 3). Importantly, no statistically significant differences were observed between intervention modalities, and improvements were largely consistent across rural and urban participants, a major covariate.

To our knowledge, this study is among the first to rigorously evaluate and directly compare different digital diabetes self-management programs using an RCT design in this way. All 3 interventions—vMMWD, TBES, and the combined approach—led to significant improvements in diabetes-related confidence, distress, and self-care behaviors, with effects through 6 months and supported by moderate to strong effect sizes. These findings generally reflect positive findings reported in other in-person and digital health diabetes interventions [8,9,16,17,21-23,43-46]. However, the TBES intervention did not result in significant improvements in diabetes-related knowledge from baseline to follow-up, contrary to other research showing that coaching-based apps can significantly increase self-management diabetes knowledge.

While all interventions demonstrated overall positive outcomes, no statistically significant superiority was observed among the intervention modalities, consistent with HbA_{1c} outcome findings reported elsewhere [6,55]. These results offer partial support for the study's hypotheses, confirming the effectiveness of each intervention but not the expected added benefit of the combined approach. This suggests the observed improvements were likely due to shared attributes (ie, DSMES components) rather than the intervention modality. Thus, whichever modality is easily accessible to participants is likely to benefit their self-care behaviors.

The study findings also partially supported our expectations that those from underserved or low-resourced settings with greater need would experience stronger positive effects, providing evidence for the heterogeneous treatment effect [56]. However, with regard to rurality, participants in both rural and urban settings experienced significant improvements across outcomes, with no statistically significant differences observed between groups. While rural participants showed comparable—if not slightly larger—gains in some outcomes (eg, knowledge and self-care), the study was not sufficiently powered to detect interaction effects. These findings suggest that digital DSMES may help mitigate geographic disparities in access to diabetes education.

The study findings can be interpreted through the lens of access, cost, and quality in diabetes self-management. From a quality perspective, all 3 interventions demonstrated encouraging effectiveness. Our prior research [6] demonstrated improvements in HbA_{1c} values, whereas this follow-up research revealed positive impacts on psychosocial factors linked to clinical outcomes [57,58]. Specifically, our DSMES-based interventions significantly improved psychological outcomes—such as confidence and distress—as well as actual self-care behaviors. These results suggest that the interventions enhanced participants' understanding and confidence in managing their diabetes, reduced diabetes-related distress, and facilitated real behavioral changes. As emphasized in classic models of chronic disease management, sustained behavioral improvements often arise from underlying psychological shifts—particularly in self-efficacy and emotional readiness [59]—which are foundational in effective chronic illness care [60].

These concepts are further supported by the Chronic Care Model [61], which underscores the importance of informed, activated patients and prepared, proactive health care teams. Prior research supports the notion that self-management programs are most effective when they improve knowledge and confidence while alleviating distress [8,41,62]. Importantly, this study also offers evidence of sustainability, with program effects maintained from baseline through 3 and 6 months. This is especially encouraging, given previously mixed findings in the literature regarding the long-term effectiveness of digital diabetes self-management programs [63,64].

The longitudinal regression findings can also be interpreted as supporting the social cognitive theory. The positive association between confidence and self-care behaviors is consistent with evidence that self-efficacy is a key driver of sustained behavior change [59,60], while the negative effect of diabetes-related distress reflects well-documented emotional barriers that undermine adherence and self-management capacity [50,62]. Together, these results reinforce that behavior change is not driven by knowledge alone, but by psychosocial mechanisms that enable individuals to translate knowledge into action. By demonstrating that confidence and distress—but not knowledge—predict longitudinal improvements in self-care, this study strengthens the theoretical understanding of how DSMES interventions influence behavior over time.

From an access perspective, this study highlights the feasibility and value of digital diabetes self-management programs. Accelerated by advancements in digital technology and the societal shifts prompted by the COVID-19 pandemic, there is increasing acceptance of accessing health care services via digital platforms [65]. Given persistent barriers to in-person care—particularly in rural areas where geographic isolation, limited transportation options, and inadequate health care infrastructure prevail—these findings support the use of digital programs as a viable alternative to traditional in-person services [66,67].

A prior meta-analysis further supports the potential of digital diabetes self-management education to improve clinical and behavioral outcomes while also highlighting significant variability in program implementation, user engagement, and

XSL•FO
RenderX

access across populations [63]. Policymakers, public health professionals, and individuals with diabetes should consider DSMES as a practical and scalable solution, provided that participants have access to the necessary devices and internet connectivity. Notably, leveraging community resources—such as public libraries with Wi-Fi and private telehealth spaces—could bridge the digital divide and extend digital health access in underserved communities [68]. From a cost perspective, self-management remains a critical pillar of diabetes care, with substantial evidence supporting both the effectiveness and cost-efficiency of DSMES [8,69]. Although this study did not directly assess the cost of digital DSMES, its findings contribute to the growing literature supporting the economic and practical value of digital delivery methods. Digital interventions can reduce or eliminate common logistical barriers—such as travel time, facility utilization, and provider burden—making them particularly advantageous for underserved and rural populations. Prior studies have shown that telehealth-based diabetes interventions significantly enhance access to care and treatment adherence in rural settings by minimizing travel burden and expanding provider reach [70-72].

Future research steps include examining the mechanisms through which digital interventions improve diabetes outcomes—particularly the interrelationships between knowledge, confidence, emotional distress, and self-care behaviors. A deeper understanding of these pathways can inform the design of more targeted, effective interventions. Ultimately, digital DSMES programs hold promise not only for improving health outcomes (quality) but also for increasing accessibility to care and reducing the economic burden associated with diabetes management—advancing the triple aim of better care, improved population health, and lower costs [73].

Limitations

This study has several limitations typical in single-site randomized clinical trial studies. First, the sample was limited to one state (Texas), which may affect the generalizability of findings to other geographic regions. However, this state includes a large and geographically diverse population (including an international border, many rural counties, and several major metropolitan areas). Furthermore, Texas is located within the South census region with a history of high rates of diabetes [2,4], with other work even pointing to higher diabetes mortality rates relative to other census regions [74]. Thus, Texas is a unique area, yet also representative of various sociodemographic groups.

Second, we were not able to recruit as many participants from rural areas as we had projected, constraining our ability to conduct formal subgroup analyses. While preimplementation sample size estimates indicated adequate power for testing main effects of DSMES programs with moderate or larger effect sizes, the relatively small sample size within each intervention arm may have limited the ability to detect differences between rural and urban subgroups. Hence, we were not able to distinguish relative intervention benefits for rural versus urban populations.

Third, all participants were required to have internet access and a smartphone, potentially introducing selection bias and limiting generalizability to those with lower digital literacy or access. However, according to a Pew Research Survey, the percentage of Americans who “rely on their smartphones for going online” (percentage of adults with a smartphone, but no broadband at home) has seen increases across all income categories from 2013 to 2021 [75]. In fact, the largest gains were for those with household incomes of less than US $30,000 (the lowest income level in the study) from 12% to 27% from 2013 to 2021 [75]. Thus, this is actually a strength, in that this method (eg, use of a smartphone to access diabetes-related training) is likely to gain more momentum over time with more Americans becoming more familiar with their smartphones as a major means of accessing information, etc. In fact, more than 90% of study participants reported being comfortable with smartphones, minimizing concerns about digital literacy in this group of participants. Yet, as indicated earlier, our study participants may not be generalizable to the broader population of people living with diabetes who may have lower digital literacy and underscore equity concerns needing policy attention [76,77].

Fourth, all outcome measures were based on self-report, which may be subject to recall or social desirability bias. However, this is not uncommon in survey research and is applicable to even large nationally representative surveys such as the BRFSS [78].

Fifth, the study did not collect data on participants’ actual engagement with the digital interventions, limiting insights into usage patterns and dose-response relationships. However, the introduction of random assignment was a major strength that likely led to greater internal validity [54].

Finally, although our study had a 6-month follow-up period revealing short-term maintenance effects, a longer follow-up period, while not feasible during our study period, would have addressed the extent to which benefits from different digital interventions could be maintained over time.

Conclusions

This RCT demonstrates the effectiveness of 3 digital DSMES interventions—vMMWD, TBES, and a combined modality—in improving diabetes-related confidence, reducing distress, and enhancing self-care behaviors among adults with T2DM. While most intervention groups showed improvements through 6 months on all diabetes-related outcomes, no single modality proved statistically superior. One notable exception emerged; knowledge gains were not significant in the TBES-only group. These findings underscore the value of digital DSMES as a flexible and scalable approach to support behavior change. Importantly, the study highlights how digital interventions can promote health equity by expanding access to self-care education for individuals in underserved and rural settings. Although the study did not directly assess costs, the digital format’s ability to reduce logistical barriers, such as travel, provider burden, and infrastructure demands, suggests strong potential for cost-efficiency.

These findings have important implications for public health practice and health care delivery. Digital DSMES programs represent a scalable and adaptable approach to supporting individuals with diabetes, particularly in underserved and resource-limited settings. Future research should explore the

XSL•FO
RenderX

mechanisms linking knowledge, confidence, and distress to behavioral outcomes to optimize digital delivery strategies. Furthermore, given recent advancements in artificial intelligence, we also expect additional improvements in the user interface (including how individuals access digital data) on smartphones with potential opportunities for subject-specific (where applicable) tailoring of knowledge and other outcomes to improve the user experience in the future. For example, digital DSMES programs and platforms could incorporate adaptive messaging that responds to users' ongoing data, such as increasingly motivational prompts and action steps when confidence scores decline [79]. Future research should focus on identifying the specific components and mechanisms that drive intervention effectiveness, including the roles of engagement, personalization, and behavioral support. Taken together, this work supports the promise of digital DSMES in advancing the goals of improved care quality, enhanced access, and reduced costs—aligned with the broader aim of transforming chronic disease management.

Acknowledgments

The authors would like to thank Drs Carrie Byington and Nancy Dickey for their leadership and support at the Texas A&M Health Science Center, with special appreciation for their vision in creating the Rural Health Moonshot Initiative. They also thank Keri Carpenter, MPH, who served as program manager for this study. The authors declare they have not used artificial intelligence tools in the creation of this paper.

Funding

This research was supported by a grant from Blue Cross Blue Shield of Texas (grant 23-468611) to establish a Rural Health Moonshot Program at Texas A&M Health Science Center. The principal investigators on this project are MGO and MLS. All claims expressed in this paper are solely those of the authors and do not necessarily represent those of their affiliated organizations. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Data Availability

The datasets presented in this paper are not readily available because the data that support the findings of this study are available from Blue Cross Blue Shield of Texas, but restrictions apply to the availability of these data, which were used under license for this study, and so are not publicly available. Requests to access the datasets should be directed to the Texas A&M point of contact for all data (Mr. Jim Colson, Texas A&M Vice President, Digital Health, jim.colson@tamu.edu).

Authors' Contributions

MO contributed to the overall conceptualization of the study, wrote the first draft of the manuscript, edited subsequent drafts, and provided funding and administrative oversight for the project (co-principal investigator). CHE contributed to the overall conceptualization of the study, wrote the first draft of the manuscript, and edited subsequent drafts. GH contributed to the overall conceptualization of the study, conducted the data analysis, and wrote the first draft of the manuscript. SDCT contributed to writing the first draft of the manuscript and editing subsequent drafts. SA contributed to writing the first draft of the manuscript and editing subsequent drafts. MLS contributed to the overall conceptualization of the study, conducted the data analysis, and provided funding and administrative oversight for the project (co-principal investigator).

Checklist 1
CONSORT checklist.
[PDF File, 73 KB - diabetes_v11i1e87364_app1.pdf]

References

1. National diabetes statistics report. The Centers for Disease Control and Prevention. 2024 May. URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2025-07-29]
2. Towne SD, Bolin J, Ferdinand A, Nicklett EJ, Smith ML, Ory MG. Assessing diabetes and factors associated with foregoing medical care among persons with diabetes: disparities facing American Indian/Alaska native, Black, Hispanic, low income, and southern adults in the U.S. (2011-2015). Int J Environ Res Public Health 2017 Apr 26;14(5):464. [doi: 10.3390/ijerph14050464] [Medline: 28445431]
3. Probst JC, Moore CG, Glover SH, Samuels ME. Person and place: the compounding effects of race/ethnicity and rurality on health. Am J Public Health 2004 Oct;94(10):1695-1703. [doi: 10.2105/ajph.94.10.1695] [Medline: 15451735]
4. Towne SD, Probst JC, Smith ML, Salinas M, Ory MG. Rural health and aging: global perspectives. In: Pachana NA, editor. Encyclopedia of Geropsychology: Springer Singapore; 2017:2098-2105. [doi: 10.1007/978-981-287-082-7_257]
5. Put the brakes on diabetes complications. The Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/prevention-type-2/stop-diabetes-complications.html [accessed 2025-08-06]

XSL•FO
RenderX

6. Ory MG, Han G, Nsobundu C, Carpenter K, Towne SD, Smith ML. Comparative effectiveness of diabetes self-management education and support intervention strategies among adults with type 2 diabetes in Texas. Front Public Health 2025;13:1543298. [doi: 10.3389/fpubh.2025.1543298] [Medline: 40171438]
7. Owolabi EO, Boakye MDS, Omololu SO, Smalls BL, Shaibi GQ. Technology-based interventions to promote diabetes self-management behaviors for persons newly diagnosed with type 2 diabetes: a scoping review. Curr Diab Rep 2024 Dec;24(12):257-272. [doi: 10.1007/s11892-024-01553-3] [Medline: 39352613]
8. Powers MA, Bardsley JK, Cypress M, et al. Diabetes self-management education and support in adults with type 2 diabetes: a consensus report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. Diabetes Care 2020 Jul;43(7):1636-1649. [doi: 10.2337/dci20-0023] [Medline: 32513817]
9. Mendez I, Lundeen EA, Saunders M, Williams A, Saaddine J, Albright A. Diabetes self-management education and association with diabetes self-care and clinical preventive care practices. Sci Diabetes Self Manag Care 2022 Feb;48(1):23-34. [doi: 10.1177/26350106211065378] [Medline: 35023406]
10. Gray KE, Hoerster KD, Taylor L, Krieger J, Nelson KM. Improvements in physical activity and some dietary behaviors in a community health worker-led diabetes self-management intervention for adults with low incomes: results from a randomized controlled trial. Transl Behav Med 2021 Dec 14;11(12):2144-2154. [doi: 10.1093/tbm/ibab113] [Medline: 34424331]
11. Ahmad F, Joshi SH. Self-care practices and their role in the control of diabetes: a narrative review. Cureus 2023 Jul;15(7):e41409. [doi: 10.7759/cureus.41409] [Medline: 37546053]
12. Hill CV, Pérez-Stable EJ, Anderson NA, Bernard MA. The National Institute on Aging health disparities research framework. Ethn Dis 2015 Aug 7;25(3):245-254. [doi: 10.18865/ed.25.3.245] [Medline: 26675362]
13. A conceptual framework for action on the social determinants of health. World Health Organization. 2010. URL: https://www.who.int/publications/i/item/9789241500852 [accessed 2025-08-08]
14. Zuercher E, Diatta ID, Burnand B, Peytremann-Bridevaux I. Health literacy and quality of care of patients with diabetes: a cross-sectional analysis. Prim Care Diabetes 2017 Jun;11(3):233-240. [doi: 10.1016/j.pcd.2017.02.003] [Medline: 28292571]
15. van der Heide I, Uiters E, Rademakers J, Struijs JN, Schuit AJ, Baan CA. Associations among health literacy, diabetes knowledge, and self-management behavior in adults with diabetes: results of a Dutch cross-sectional study. J Health Commun 2014;19 Suppl 2(sup2):115-131. [doi: 10.1080/10810730.2014.936989] [Medline: 25315588]
16. Camargo-Plazas P, Robertson M, Alvarado B, Paré GC, Costa IG, Duhn L. Diabetes self-management education (DSME) for older persons in Western countries: a scoping review. In: Mendoza-Nuñez VM, editor. PLoS ONE 2023;18(8):e0288797. [doi: 10.1371/journal.pone.0288797] [Medline: 37556399]
17. Peña-Purcell N, Han G, Lee Smith M, Peterson R, Ory MG. Impact of diabetes self-management education on psychological distress and health outcomes among African Americans and Hispanics/Latinos with diabetes. Diabetes Spectr 2019 Nov;32(4):368-377. [doi: 10.2337/ds18-0081] [Medline: 31798295]
18. Healthy days methods and measures. The Centers for Disease Control and Prevention. 2018. URL: https://archive.cdc.gov/www_cdc_gov/hrqol/methods.htm [accessed 2025-08-08]
19. Bober T, Rothenberger S, Lin J, Ng JM, Zupa M. Factors associated with receipt of diabetes self-management education and support for type 2 diabetes: potential for a population health management approach. J Diabetes Sci Technol 2023 Sep;17(5):1198-1205. [doi: 10.1177/19322968231176303] [Medline: 37264614]
20. Xie Z, Liu K, Or C, Chen J, Yan M, Wang H. An examination of the socio-demographic correlates of patient adherence to self-management behaviors and the mediating roles of health attitudes and self-efficacy among patients with coexisting type 2 diabetes and hypertension. BMC Public Health 2020 Aug 12;20(1):1227. [doi: 10.1186/s12889-020-09274-4] [Medline: 32787809]
21. Siminerio L, Krall J, Johnson P, et al. Examining a diabetes self-management education and support telemedicine model with high-risk patients in a rural community. J Diabetes Sci Technol 2023 Sep;17(5):1190-1197. [doi: 10.1177/19322968231180884] [Medline: 37338130]
22. Nkhoma DE, Soko CJ, Bowrin P, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021 Oct;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
23. Enricho Nkhoma D, Jenya Soko C, Joseph Banda K, Greenfield D, Li YCJ, Iqbal U. Impact of DSMES app interventions on medication adherence in type 2 diabetes mellitus: systematic review and meta-analysis. BMJ Health Care Inform 2021 Apr;28(1):e100291. [doi: 10.1136/bmjhci-2020-100291] [Medline: 33853862]
24. Rogers EM. Diffusion of Innovations, 4th edition: Free Press; 1995.
25. Yudkin JS, Messiah SE, Allicock MA, Barlow SE. Integration of e-Health strategies for post-COVID-19 pandemic pediatric weight management programs. Telemed J E Health 2024 Feb;30(2):321-330. [doi: 10.1089/tmj.2023.0068] [Medline: 37552819]

XSL•FO
RenderX

26. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024 Jul 22;26:e55757. [doi: 10.2196/55757] [Medline: 39037772]
27. Horigan G, Davies M, Findlay-White F, Chaney D, Coates V. Reasons why patients referred to diabetes education programmes choose not to attend: a systematic review. Diabet Med 2017 Jan;34(1):14-26 [FREE Full text] [doi: 10.1111/dme.13120] [Medline: 26996982]
28. Markowitz B, de Sequeira S, Najam A, et al. Beyond hemoglobin A1c-outcomes that matter to individuals with type 1 diabetes in adopting digital health interventions for self-management support: qualitative study. JMIR Diabetes 2024 Nov 7;9:e60190. [doi: 10.2196/60190] [Medline: 39509700]
29. Jang Y, Yang Y. Effects of e-health literacy on health-related quality of life in young adults with type 2 diabetes: parallel mediation of diabetes self-efficacy and self-care behaviors. Appl Nurs Res 2025 Apr;82:151917. [doi: 10.1016/j.apnr.2025.151917] [Medline: 40086937]
30. Tajdar D, Lühmann D, Walther L, Bittner L, Scherer M, Schäfer I. Effects of two COVID-19 lockdowns on HbA1c levels in patients with type 1 diabetes and associations with digital treatment, health literacy, and diabetes self-management: a multicenter, observational cohort study over 3 years. Diabetes Ther 2024 Jun;15(6):1375-1388. [doi: 10.1007/s13300-024-01574-x] [Medline: 38642263]
31. Nevoret C, Gervaise N, Delemer B, et al. The effectiveness of an app (Insulia) in recommending basal insulin doses for French patients with type 2 diabetes mellitus: longitudinal observational study. JMIR Diabetes 2023 Mar 1;8:e44277. [doi: 10.2196/44277] [Medline: 36749650]
32. Patnaik L, Panigrahi SK, Sahoo AK, Mishra D, Beura S, Muduli AK. Mobile health application based intervention for improvement of quality of life among newly diagnosed type 2 diabetes patients. Clinical Diabetology . [doi: 10.5603/DK.a2021.0014]
33. Omar MA, Hasan S, Palaian S, Mahameed S. The impact of a self-management educational program coordinated through WhatsApp on diabetes control. Pharm Pract (Granada) 2020;18(2):1841. [doi: 10.18549/PharmPract.2020.2.1841] [Medline: 32477434]
34. Kelley H, Chiasson M, Downey A, et al. The clinical impact of eHealth on the self-management of diabetes: a double adoption perspective. JAIS 2011 Mar;12(3):208-234. [doi: 10.17705/1jais.00263]
35. Song M, Choe MA, Kim KS, et al. An evaluation of web-based education as an alternative to group lectures for diabetes self-management. Nurs Health Sci 2009 Sep;11(3):277-284 [FREE Full text] [doi: 10.1111/j.1442-2018.2009.00458.x] [Medline: 19689636]
36. Sinha B, Ghosal S. A target HbA1c between 7 and 7.7% reduces microvascular and macrovascular events in T2D regardless of duration of diabetes: a meta-analysis of randomized controlled trials. Diabetes Ther 2021 Jun;12(6):1661-1676. [doi: 10.1007/s13300-021-01062-6] [Medline: 33895981]
37. American Diabetes Association Professional Practice Committee. 6. Glycemic targets: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 1;45(Suppl 1):S83-S96. [doi: 10.2337/dc22-S006] [Medline: 34964868]
38. Ory MG, Lee S, Towne SD, Flores S, Gabriel O, Smith ML. Implementing a diabetes education program to reduce health disparities in South Texas: application of the RE-AIM framework for planning and evaluation. Int J Environ Res Public Health 2020 Aug 30;17(17):32872662. [doi: 10.3390/ijerph17176312] [Medline: 32872662]
39. Smith ML, Zhong L, Lee S, Towne SD, Ory MG. Effectiveness and economic impact of a diabetes education program among adults with type 2 diabetes in South Texas. BMC Public Health 2021 Sep 9;21(1):1646. [doi: 10.1186/s12889-021-11632-9] [Medline: 34503468]
40. Davis J, Fischl AH, Beck J, et al. 2022 national standards for diabetes self-management education and support. Sci Diabetes Self Manag Care 2022 Feb;48(1):44-59. [doi: 10.1177/26350106211072203] [Medline: 35049403]
41. Funnell MM, Brown TL, Childs BP, et al. National standards for diabetes self-management education. Diabetes Care 2010 Jan;33 Suppl 1(Suppl 1):S89-S96. [doi: 10.2337/dc10-S089] [Medline: 20042780]
42. Alexandre K, Campbell J, Bugnon M, et al. Factors influencing diabetes self-management in adults: an umbrella review of systematic reviews. JBI Evid Synth 2021 May;19(5):1003-1118. [doi: 10.11124/JBIES-20-00020] [Medline: 33741836]
43. Wilmoth SR, Carrillo-McCracken LL, Wilhite B, et al. Impacts of diabetes self-management education and support programs in Hispanic church settings: a cluster-randomized trial comparing faith-based and faith-placed approaches. Nutrients 2024 Dec 28;17(1):69. [doi: 10.3390/nu17010069] [Medline: 39796503]
44. Dietz CJ, Sherrill WW, Ankomah S, Rennert L, Parisi M, Stancil M. Impact of a community-based diabetes self-management support program on adult self-care behaviors. Health Educ Res 2023 Jan 20;38(1):1-12. [doi: 10.1093/her/cyac034] [Medline: 36367205]
45. Ernawati U, Wihastuti TA, Utami YW. Effectiveness of diabetes self-management education (DSME) in type 2 diabetes mellitus (T2DM) patients: systematic literature review. J Public Health Res 2021 Apr 14;10(2):2240. [doi: 10.4081/jphr.2021.2240] [Medline: 33855427]
46. Okeyo HM, Biddle M, Williams LB. Impact of diabetes self-management education on A1C levels among Black/African Americans: a systematic review. Sci Diabetes Self Manag Care 2024 Feb;50(1):87-95. [doi: 10.1177/26350106231213400] [Medline: 37994435]

XSL•FO
RenderX

47. Rothman RL, Malone R, Bryant B, et al. The spoken knowledge in low literacy in diabetes scale. Diabetes Educ 2005 Mar;31(2):215-224. [doi: 10.1177/0145721705275002] [Medline: 15797850]
48. Peña-Purcell NC, Lee J, Ory MG. Integrating a Care Supporter in Diabetes Self-Management Education to Enhance Health Outcomes: A Pilot Study: The Science of Diabetes Self-Management and Care SAGE Publications; 2025. [doi: 10.1177/26350106251320327]
49. Self-Management Resource Center. URL: https://selfmanagementresource.com/ [accessed 2025-08-06]
50. Polonsky WH, Fisher L, Earles J, et al. Assessing psychosocial distress in diabetes. Diabetes Care 2005 Mar 1;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
51. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
52. NCHS urban-rural classification scheme for counties. CDC. 2025. URL: https://www.cdc.gov/nchs/data-analysis-tools/urban-rural.html [accessed 2025-08-06]
53. Fayers PM, Sprangers MAG. Understanding self-rated health. Lancet 2002 Jan 19;359(9302):187-188 [FREE Full text] [doi: 10.1016/S0140-6736(02)07466-4] [Medline: 11812551]
54. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLoS Clin Trials 2006 May;1(1):e9. [doi: 10.1371/journal.pctr.0010009] [Medline: 16871331]
55. Jafar N, Huriyati E, Setyawati A. Enhancing knowledge of diabetes self-management and quality of life in people with diabetes mellitus by using Guru Diabetes apps-based health coaching. J Public Health Res 2023 Jul;12(3):22799036231186338. [doi: 10.1177/22799036231186338] [Medline: 37465530]
56. Hu A. Heterogeneous treatment effects analysis for social scientists: a review. Soc Sci Res 2023 Jan;109:102810. [doi: 10.1016/j.ssresearch.2022.102810] [Medline: 36470639]
57. Spruijt-Metz D, O'Reilly GA, Cook L, Page KA, Quinn C. Behavioral contributions to the pathogenesis of type 2 diabetes. Curr Diab Rep 2014 Apr;14(4):475. [doi: 10.1007/s11892-014-0475-3] [Medline: 24604714]
58. Chew BH, Shariff-Ghazali S, Fernandez A. Psychological aspects of diabetes care: effecting behavioral change in patients. World J Diabetes 2014 Dec 15;5(6):796-808. [doi: 10.4239/wjd.v5.i6.796] [Medline: 25512782]
59. Bandura A. Self-Efficacy: The Exercise of Control: W.H. Freeman; 1997.
60. Glasgow RE, Funnell MM, Bonomi AE, Davis C, Beckham V, Wagner EH. Self-management aspects of the improving chronic illness care breakthrough series: implementation with diabetes and heart failure teams. Ann Behav Med 2002;24(2):80-87. [doi: 10.1207/S15324796ABM2402_04] [Medline: 12054323]
61. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
62. Fisher L, Glasgow RE, Mullan JT, Skaff MM, Polonsky WH. Development of a brief diabetes distress screening instrument. Ann Fam Med 2008;6(3):246-252. [doi: 10.1370/afm.842] [Medline: 18474888]
63. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
64. Pal K, Eastwood SV, Michie S, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013 Mar 28;2013(3):CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
65. Shaw RJ, Bonnet JP, Modarai F, George A, Shahsahebi M. Mobile health technology for personalized primary care medicine. Am J Med 2015 Jun;128(6):555-557. [doi: 10.1016/j.amjmed.2015.01.005] [Medline: 25613298]
66. Weinhold I, Gurtner S. Understanding shortages of sufficient health care in rural areas. Health Policy 2014 Nov;118(2):201-214. [doi: 10.1016/j.healthpol.2014.07.018] [Medline: 25176511]
67. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Public Health (Fairfax) 2015 Jun;129(6):611-620. [doi: 10.1016/j.puhe.2015.04.001] [Medline: 26025176]
68. Spaulding EM, Miller HN, Metlock FE, et al. Leveraging community Wi-Fi and spaces for digital health use. Front Public Health 2024;12:1418627. [doi: 10.3389/fpubh.2024.1418627] [Medline: 38912273]
69. Duncan I, Birkmeyer C, Coughlin S, Li QE, Sherr D, Boren S. Assessing the value of diabetes education. Diabetes Educ 2009;35(5):752-760. [doi: 10.1177/0145721709343609] [Medline: 19783766]
70. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 1;27(6):939-945. [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
71. Izquierdo RE, Knudson PE, Meyer S, Kearns J, Ploutz-Snyder R, Weinstock RS. A comparison of diabetes education administered through telemedicine versus in person. Diabetes Care 2003 Apr;26(4):1002-1007. [doi: 10.2337/diacare.26.4.1002] [Medline: 12663564]
72. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495. [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]

XSL•FO
RenderX

73. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
74. Callaghan T, Ferdinand AO, Akinlotan MA, Towne SD, Bolin J. The changing landscape of diabetes mortality in the United States across region and rurality, 1999-2016. J Rural Health 2020 Jun;36(3):410-415. [doi: 10.1111/jrh.12354] [Medline: 30802321]
75. Vogels EA. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2021. URL: https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2026-05-07]
76. Nagori A, Keshvani N, Patel L, Dhruve R, Sumarsono A. Electronic health literacy gaps among adults with diabetes in the United States: role of socioeconomic and demographic factors. Prev Med Rep 2024 Nov;47:102895. [doi: 10.1016/j.pmedr.2024.102895] [Medline: 39823076]
77. Sabben G, Telfort C, Morales M, et al. Technology and continuous glucose monitoring access, literacy, and use among patients at the diabetes center of an inner-city safety-net hospital: mixed methods study. JMIR Diabetes 2024 Oct 15;9:e54223. [doi: 10.2196/54223] [Medline: 39405528]
78. Mokdad AH, Stroup DF, Giles WH, Behavioral Risk Factor Surveillance Team. Public health surveillance for behavioral risk factors in a changing environment. Recommendations from the behavioral risk factor surveillance team. MMWR Recomm Rep 2003 May 23;52(RR-9):1-12. [Medline: 12817947]
79. Celano CM, Massey C, Long J, et al. An adaptive, algorithm-based text message intervention to promote health behavior adherence in type 2 diabetes: treatment development and proof-of-concept trial. J Diabetes Sci Technol 2023 Mar;17(2):364-373. [doi: 10.1177/19322968211065067] [Medline: 34911398]

Abbreviations:

CONSORT: Consolidated Standards of Reporting Trials
DSMES: Diabetes Self-Management Education and Support
HbA_{1c}: hemoglobin A_{1c}
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus
TBES: Technology-Based Education and Support
vMMWD: Virtual Making Moves with Diabetes

Edited by I Steenstra; submitted 07.Nov.2025; peer-reviewed by B Bock, J Dai; revised version received 07.Apr.2026; accepted 07.Apr.2026; published 01.Jun.2026.

Please cite as:
Ory M, Esquivel CH, Han G, Towne Jr SDC, Ahn S, Smith ML
Personalized Digital Health Solutions to Increasing Diabetes-Related Knowledge and Behavioral Outcomes: Results From a Randomized Controlled Trial
JMIR Diabetes 2026;11:e87364
URL: https://diabetes.jmir.org/2026/1/e87364
doi:10.2196/87364

© Marcia Ory, Christi H Esquivel, Gang Han, Samuel D Castiglione Towne Jr, SangNam Ahn, Matthew Lee Smith. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 1.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Promoting Diabetes Self-Management Among Vietnamese Americans: Mixed Methods Pilot Study

Anna Nguyen[1], MSN, PhD; Valerie Eschiti[1], MSN, PhD; Thanh C Bui[1], MD, DrPH; Katherine O'Neal[1], PharmD; Tam Nguyen[2], MSN, MSPH, PhD; Angelina P Nguyen[3], PhD; Hanxia Li[1], MS; Michael Machiorlatti[1], PhD; Kathleen Dwyer[1], PhD

[1]University of Oklahoma Health Campus, 1100 N. Stonewall Avenue, Oklahoma City, OK, United States

[2]Boston College, Chestnut Hill, MA, United States

[3]Baylor University, Dallas, TX, United States

Corresponding Author:
Anna Nguyen, MSN, PhD
University of Oklahoma Health Campus, 1100 N. Stonewall Avenue, Oklahoma City, OK, United States

Abstract

Background: Participating in a Diabetes Self-Management Education and Support (DSMES) program improves self-care behaviors, quality of life, and health outcomes. However, language barriers and cultural differences can hinder participation, leaving many Vietnamese Americans with limited access to DSMES services.

Objective: This study aims to evaluate the feasibility, acceptability, and preliminary efficacy of a 3-month Blended Automated Links Augmented by Nurse Call and Engagement (BALANCE) intervention designed to deliver culturally tailored DSMES in the Vietnamese language, with participants monitored for 12 months afterward to assess sustained effects on key outcomes.

Methods: An explanatory sequential mixed methods design was used, guided by the Practical, Robust Implementation and Sustainability Model (PRISM) framework. Feasibility and acceptability were measured by the participation rate of eligible clinics and patients, patient message response rate, and retention rate. Focus groups were conducted to assess adoption and sustainability. A pilot single-arm, prospective interventional trial was conducted with a sample of 88 Vietnamese American adults with type 2 diabetes from 10 primary care clinics. Surveys were administered at baseline and every 3 months over 12 months. Repeated measures ANOVA assessed changes in clinical outcomes at 3, 6, 9, and 12 months. Qualitative data from in-depth interviews and focus groups were thematically analyzed to validate and expand on quantitative findings. Integrated analysis using joint display enabled meta-inferences across data sources.

Results: Among 88 participants (mean age 68, SD 9.8; range 35‑86 years), the intervention did not significantly affect glycated hemoglobin A_{1c} (P=.63) but led to a statistically and clinically significant reduction in low-density lipoprotein (P=.001) and improvement in exercise performance (P=.04). Qualitative data from 45 patient interviews reached data saturation, with 80% (n=36) describing the intervention as “convenient” and “helpful.” Clinic staff (n=18) participated in 3 focus groups and endorsed the intervention as acceptable and feasible. Mixed methods analysis confirmed high feasibility (83% clinic participation and 100% clinic retention) and acceptability (90.9% patient retention). Key barriers to sustainability included limited staffing and supply infrastructure.

Conclusions: Intervention feasibility and acceptability were demonstrated but require further refinement to achieve long-term, consistent glycemic control. Findings indicated that clinic staff workload and clinic workflow were key determinants of the study’s feasibility and acceptability. Future research should test BALANCE in a fully powered randomized controlled trial to evaluate intervention effectiveness.

(JMIR Diabetes 2026;11:e80177) doi:10.2196/80177

KEYWORDS

Vietnamese Americans; diabetes self-management; mobile health; mixed methods; nursing

Introduction

Background

In the United States, 37.3 million people (11.3% of the population) have diabetes, with 8.25 million diabetes-related hospitalizations reported annually (327.9/1000) [1]. In 2022, the cost of diabetes in the United States was US $412.9 billion, including US $306.6 billion in direct medical costs and US $106.3 billion in indirect costs. Major contributors to indirect costs include reduced employment due to disability and lost productivity from premature deaths [2]. Diabetes is a

XSL•FO
RenderX

self-managed disease, and effective management can reduce the risk of complications, morbidity, and health care costs [3,4]. Evidence indicates that participating in a Diabetes Self-Management Education and Support (DSMES) program can reduce the risk of diabetes complications and improve health outcomes [5,6].

While diabetes is the seventh leading cause of death in the overall US population, it ranks as the fifth leading cause of death among Asian Americans [7]. As members of the Asian American group, Vietnamese Americans (VNAs) face barriers to participating in existing DSMES programs due to language and cultural differences, which hinder their ability to obtain appropriate self-management information [8]. There is a limited number of Vietnamese-speaking diabetes care and education specialists, and the DSMES resources in the Vietnamese language are scarce. Language barriers often prevent effective communication with health care providers, and differences in cultural beliefs and practices may not align with standard DSMES materials. The Vietnamese community has recognized these barriers and expressed a need for formal diabetes education and support in the Vietnamese language [9].

VNAs are the fourth largest Asian American group with 90.6% of persons diagnosed with diabetes falling within the nonobese BMI range of less than 30. Nonobese VNAs have 60% higher adjusted odds of diabetes compared to nonobese non-Hispanic Whites [10]. Given the high prevalence of diabetes among VNAs, ensuring access to DSMES services is critical. The 2022 National Standards for DSMES highlight the importance of delivering services that respect cultural diversity, address social determinants of health, and leverage modern technology for engagement [5]. With mobile phone usage now nearly universal—93% of the global population has access to a mobile broadband network—mobile health (mHealth) solutions offer a powerful and accessible platform for delivering DSMES services [11].

Although DSMES programs have demonstrated effectiveness, the evidence base for interventions tailored specifically to VNAs remains limited. Prior research has primarily focused on (1) documenting culturally influenced beliefs and practices that shape diabetes self-management and help-seeking behaviors, highlighting the importance of aligning interventions with Vietnamese health beliefs and trusted information sources, and (2) emphasizing the value of community-engaged approaches to culturally tailoring DSMES content and delivery. Building on this emerging body of work, this study is novel in its implementation of a culturally adapted mHealth DSMES intervention designed to address language discordance and limited access to culturally relevant education in supporting VNAs with diabetes.

Evidence suggests that incorporating mHealth technology to provide education and support lifestyle and health behavior changes is feasible and acceptable for various populations [12-15], including Vietnamese individuals living in Vietnam [16]. However, current literature lacks evidence on the feasibility, acceptability, and effectiveness of culturally adapted interventions that leverage mHealth technology to educate and support diabetes self-management among VNAs. Hence, this study team developed a Blended Automated Links Augmented by Nurse Call and Engagement (BALANCE) intervention using mHealth technology to deliver DSMES in the Vietnamese language.

BALANCE Intervention

BALANCE is a 3-month intervention program delivering DSMES through smartphones to empower patients in managing their diabetes by providing convenient access to knowledge, skills, and support for self-care. Following the intervention, participants were monitored for 12 months to assess sustained effects on key outcomes. The study team met with the health care providers and clinic staff at each site to discuss the referral process and operations prior to implementation. Once the clinics were ready to begin the intervention, patient education materials were distributed, and training sessions on mHealth navigation were conducted with clinic staff to prepare them for troubleshooting any patient issues with messaging. All patients received the same multicomponent intervention: (1) a printed and translated education handbook; (2) mHealth with 2-way text messaging; and (3) weekly coaching phone calls from a nurse trained in motivational interviewing techniques. These phone calls aim to reinforce the educational content, address individual challenges, and motivate patients to adhere to recommended self-care behaviors. The purpose of this study was to (1) evaluate the acceptability, feasibility, and preliminary efficacy of BALANCE and (2) assess potential factors related to BALANCE program adoption and sustainability among participating primary care clinics.

Methods

Study Design

This study used an explanatory sequential mixed methods research design to integrate findings of quantitative and qualitative data and to interpret results [17]. First, a pilot single-arm prospective interventional trial was conducted to assess BALANCE's feasibility and preliminary efficacy. This phase involved collecting and analyzing data to identify intervention reach and adoption. Second, qualitative data were collected using in-depth interviews and focus groups to explore participants' experiences and perceptions on BALANCE adoption and sustainment. Sequentially combining these methods allowed for data to be triangulated, enhanced the validity of study findings, and offered insights into the intervention's overall impact.

Inclusion and Exclusion Criteria

Clinics were eligible for inclusion if they served more than 10 VNA adults aged 18 years or older with a documented diagnosis of type 2 diabetes (T2D), and if clinicians and staff agreed to participate in BALANCE intervention training and anticipated completion of all study activities throughout the study period. At the patient level, individuals were included if they were 18 years of age or older; had a clinical diagnosis of T2D; were currently prescribed oral and/or injectable diabetes medications; owned a mobile phone capable of receiving text messages; and were able to read and write Vietnamese sufficiently to understand study materials and intervention content. Patients

XSL•FO
RenderX

were also required to demonstrate the cognitive capacity to understand study procedures and provide informed consent. Individuals were excluded if they had moderate-to-severe cognitive impairment that limited consent capacity, were currently pregnant due to differing diabetes management requirements during gestation, lacked adequate Vietnamese literacy, or did not have reliable access to a mobile device.

Measurements

The Practical, Robust Implementation and Sustainability Model (PRISM) was used to assess the outcomes and related contextual and environmental factors [18,19]. PRISM was developed from the Reach, Efficacy, Adoption, Implementation, Maintenance (RE-AIM) framework and hence includes all of the RE-AIM constructs and outcomes. PRISM is better because it takes into account external factors to provide a more comprehensive understanding of contextual influences on the implementation. The RE-AIM framework was applied with specific attention to patient-level reach and effectiveness, whereas the PRISM framework was used to explain the role of organizational and environmental factors in shaping clinic-level implementation and sustainability [19-23]. The types of data and data sources are summarized in Table 1.

XSL•FO
RenderX

Table . Assessments using the PRISM[a] framework.

RE-AIM[b] or PRISM domain	Baseline	3-month postintervention	Follow-ups	
			6 months	12 months
Patient demographics[c]	✓			
Clinic demographics[c]	✓			
Reach: absolute number of those participating in BALANCE[d]				
% of eligible clinics and patients who participated[c]	✓			
Barriers and facilitators to reach[e]	✓			
Efficacy: changes in self-care behavior, clinical laboratory results				
Changes in SDSCA[f] scale[c]	✓	✓	✓	✓
Changes in HbA_{1c}[g], triglyceride, LDL[h], and HDL[c,i]	✓	✓	✓	✓
Adoption: patient level				
Response rates to text messaging[c]	✓		✓	
% of patients enrolled at end of study period[c]	✓		✓	✓
Adoption and implementation: clinic level				
Fidelity to the study protocol and guidelines[c]		✓	✓	✓
Barriers and facilitators to implementation[e]			✓	
Sustainability: organizational infrastructure				
Sustainability of intervention at clinics[e]			✓	
Participation experiences[e]			✓	✓

[a]PRISM: Practical, Robust Implementation and Sustainability Model.
[b]RE-AIM: Reach, Efficacy, Adoption, Implementation, Maintenance.
[c]Programmatic data.
[d]BALANCE: Blended Automated Links Augmented by Nurse Call and Engagement.
[e]Interviews and focus groups.
[f]SDSCA: Summary of Diabetes Self-Care Activities.
[g]HbA_{1c}: hemoglobin A_{1c}.
[h]LDL: low-density lipoprotein.
[i]HDL: high-density lipoprotein.

Quantitative Survey and Clinical Outcomes

Routine programmatic data were collected to assess feasibility as well as clinical outcomes such as laboratory values and self-care behaviors. Reach was assessed using quantitative measures, including the recruitment and enrollment rates of clinics and their patients. In this study, reach was defined as the extent to which BALANCE engages with both clinics and patients. At the clinic level, reach was measured by the proportion of eligible clinics that agreed to participate in the study. At the patient level, reach was assessed by the proportion of eligible patients who enrolled in the study, relative to the total number of eligible individuals within participating clinics. To evaluate potential intervention efficacy, self-reported care behavior surveys and clinical laboratory test results were obtained at baseline, 3 months postintervention, and at 6 and 12 months. Self-care behavior changes were assessed using the Summary of Diabetes Self-Care Activities (SDSCA) scale, a

XSL•FO
RenderX

translated and validated tool that measures the frequency of diabetes self-care activities over the past 7 days. The tool consists of 11 items to be rated on an 8-point Likert scale (0 - 7) [24]. A higher score indicates better adherence to diet, physical activity, glucose monitoring, foot care, and medication regimen. Clinical laboratory test results were obtained from the patients' medical records to assess the intervention's impact on glycemic control and cardiovascular health, including hemoglobin A_{1c} (HbA_{1c}), low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides. Adoption was assessed by examining the acceptance and use of BALANCE intervention at both the clinic and patient levels. Implementation was assessed by looking at the fidelity of the intervention delivery, including adherence to the scheduled messages and adherence to the nurse coaching manual. In summary, 2 quantitative data sources were used, which include routine programmatic and clinical data and patient survey data. Programmatic and clinical data assessed feasibility and clinical outcomes, while survey data captured patient-reported self-care behavior outcomes.

Qualitative Interviews and Focus Groups

All 88 patients were invited to participate in an in-depth interview after completing the intervention. The health care providers of participating clinics were invited to participate in a separate focus group. All interviews and focus groups were conducted face-to-face with a member of the research team and a research assistant. Patient interviews were conducted in Vietnamese, and focus groups with clinic members were conducted in English. Data saturation was achieved following the completion of 45 patient interviews, with 5 to 10 patients from each clinic. Saturation was determined when no new themes, insights, or patterns emerged from successive interviews, indicating that additional data collection was unlikely to yield novel information relevant to the study objectives [25]. Three focus groups were conducted with 6 physicians, 1 physician assistant, and 11 clinic staff (1 to 3 per clinic). Interviews and focus groups were audio-taped, and a trained research assistant took field notes as a summary. The in-depth interview questions and focus group questions were designed to gather information related to their participation experiences.

Data Analysis

Quantitative Data

Descriptive statistics were used to summarize baseline and follow-up measurements for clinical laboratory tests and self-care behavior surveys. Repeated measures 1-way ANOVA was used to explore treatment effects at 3-, 6-, 9-, and 12-month postintervention. All analyses were conducted using Python (version 3.11.13; Python Software Foundation). A 2-sided P value of <.05 was used as the threshold for statistical significance.

Qualitative Data

Patient interview recordings were translated from Vietnamese to English before being transcribed. All recordings were transcribed and uploaded to NVivo 15 software program (QSR International) for organizing and categorizing themes. Direct content analysis was used to assess technical implementation, participant engagement, cultural relevance, language appropriateness, and BALANCE's impact on health behaviors and clinical outcomes. Direct content analysis is a structured method used to identify key concepts and is appropriate to use for analyzing predefined categories or themes [26]. This method provided valuable insights into feasibility and acceptability, as well as information that can inform the refinement and optimization of future intervention updates. Transcripts were independently coded by 2 study investigators, with differences in coding resolved via discussion. The 2 coders then identified themes and relationships among finalized codes.

Ethical Considerations

Eligible individuals went through an informed consent process to receive comprehensive information about the study, including its purpose, procedures, potential benefits, and risks. During this session, a member of the study team explained the study in detail, answered any questions, and ensured that those who are eligible fully understood what their involvement would entail. Confidentiality and data protection were also explained. Participants were assured that their personal information would be kept secure and used only for research purposes. Those interested in participating provided a written consent to participate. Before any data collection began, the study received approval from the University of Oklahoma Health Sciences Institutional Review Board (IRB 14700), ensuring that all ethical standards and guidelines were met. Study participants received US $50 in the form of a gift card as compensation for their time and effort at each of the data collecting points.

Results

Demographics and Baseline Characteristics

Clinic Level

Ten primary care clinics participated in the study. Among these, 4 clinics were independently owned and operated by a single physician, while the remaining 6 were part of larger health systems with multiple providers. All clinics, with the exception of one that accepted only private insurance, offered services to patients covered by Medicare, Medicaid, and private insurance plans. The collaborative network with primary care providers played a crucial role in integrating BALANCE into routine health care delivery. Their participation facilitated a seamless incorporation of the intervention into existing clinical workflows, supporting efficient implementation and enhancing the potential for broader adoption across varied primary care environments.

Patient Level

The study enrolled 88 adults with T2D from primary care clinics in the Central United States. Participants ranged in age from 35 to 86 years, with an average age of 68 (SD 9.8) years. Only 5% of participants (4/88) had ever attended at least one DSMES session since the time of their diagnosis, and none had completed the program. Regarding health behaviors and comorbid conditions, 5% (4/88) were current smokers, 70% (62/88) were taking hypertension medication, and 77% (68/88) were taking medication for hyperlipidemia. Within the first week of the intervention, 5% (4/88) of participants reported difficulty reading text messages or printed materials due to vision

XSL•FO
RenderX

limitations. All 4 were given the option to remain in the study, as they had live-in family members who could assist them. Refer to Table 2 for a detailed breakdown of patient demographic characteristics.

Table . Demographic characteristics of patients.

Characteristics	Total
Gender, n (%)	
Men	52 (59)
Women	36 (41)
Years with diabetes, n (%)	
0 - 1	8 (9)
2 - 3	12 (14)
4 - 5	12 (14)
6 - 7	9 (10)
8 - 9	5 (6)
>9	42 (48)
Attended some DSMES[a] since diagnosis, n (%)	
Yes	4 (5)
No	84 (96)
Years residing in the United States, n (%)	
0 - 5	8 (9)
6 - 10	4 (5)
>10	76 (86)
Received family support for diabetes care, n (%)	
Yes	39 (44)
No	49 (56)

[a]DSMES: Diabetes Self-Management Education and Support.

Feasibility and Acceptability Outcomes

Intervention Reach

To contextualize intervention reach, the clinic-level recruitment and enrollment are considered indicators of strong engagement with the primary care practice clinicians. Of the 10 eligible clinics identified, 10 (100%) agreed to participate in the study. Across these clinics, 175 eligible patients were identified and were invited to enroll. Of those invited, 88 consented to participate, resulting in a 50% patient enrollment rate. This level of participation demonstrates moderate patient-level reach, indicating feasibility and general receptiveness to the research intervention study.

Adoption and Implementation at the Clinic Level

The 3 focus groups comprised 60% males and represented various clinic roles, including physicians (n=6), physician assistants (n=1), registered nurses/clinic managers (n=2), licensed practical nurses (n=4), and medical assistants (n=5). Clinicians emphasized the importance of cultural tailoring and translating materials into Vietnamese, noting that these aspects significantly enhanced patient comprehension of their treatment regimen.

> *Providing education in the Vietnamese language is crucial given the unique cultural and linguistic barriers that many Vietnamese patients face in managing type 2 diabetes.* [Physician, P03]

> *Many patients struggle with understanding the complexities of diabetes management due to a lack of linguistically appropriate resources.* [Physician, P04]

> *This is a much-needed intervention to improve health outcomes and bridge the access gap.* [Clinic staff member, RN1]

The integration of mHealth technologies was praised for its relevance in meeting the needs of those without transportation. Physicians reported that their patients found the 2-way text messaging system and weekly nurse coaching calls beneficial for answering questions related to daily self-care activities. Additionally, motivational interviewing techniques used during nurse coaching calls were seen as particularly valuable in fostering behavior change. Three physicians (P01, P02, and P06) noted that the coaching calls added a personalized element that encouraged patients to be more engaged in their care. They noted that patients were more likely to set realistic goals and take actionable steps toward improving their diabetes management.

XSL•FO
RenderX

Adoption and Sustainment at the Patient Level

Among study participants, 89.7% (79/88) engaged with at least 2 of the 3 BALANCE intervention components. All individuals who participated in the intervention completed the postintervention self-care behavior survey. Additionally, 74% (65/88) attended one or more scheduled nurse coaching calls, indicating a strong interest in personalized support. These qualitative data were derived from the patient survey dataset and metrics extracted from the mHealth program database. Qualitative feedback from in-depth interviews with patients (n=45) highlights the acceptability, revealing high satisfaction with the intervention, particularly the ease of navigating the mHealth messages. Patients reported feeling more empowered and confident in managing their diabetes and appreciated the culturally relevant education materials. All patients valued the convenience of receiving education and support through their mobile phones, eliminating the need to travel to in-person classes, which contributed to their decision to complete the study.

> *I like the way messages are sent on a daily basis, just short and simple to understand. I also like that we can read more details in the book. And with it being in Vietnamese and showing Vietnamese food was exactly what we needed.* [Female, 65 years, participant]

> *I don't have time to drive to class and sit there for hours. Receiving text messages on my phone made it easy to learn and be informed, even with my busy schedule. This made my involvement easier.* [Female, 67 years, participant]

An analysis of weekly response rates to mHealth messaging revealed a gradual decline in engagement over the 3-month period. Patients were classified as low users if they responded to 2 or fewer messages per week, and as high users if they responded to 3 or more. In week 1, there were 24 low users, increasing steadily to 40 by week 12. These trends indicate a shift toward lower engagement, with fewer patients sustaining high levels of message interaction over time.

Through individual interviews, some patients expressed a desire for more interactive features within the intervention, such as quiz questions, to further enhance their engagement and learning experience. Others expressed that mHealth text messaging was unsuitable due to their limited data phone plans. The daily text messages and videos required more data than provided on some patients' phone plans have available, potentially leading to extra charges if data limits were exceeded and making the intervention financially burdensome. Conversely, a few patients suggested extending the duration of the intervention to provide ongoing support and reinforcement.

Organizational Infrastructure for Intervention Sustainability

Through focus groups, clinic members were asked to reflect on their experiences related to implementation and the need for sustaining BALANCE over time. A consistent theme in all focus groups was the need to integrate the DSMES service into their routine workflow, including streamlining processes to reduce administrative burdens and decrease the workload for clinic staff.

> *We see the value in this program, but to keep it going, it has to be woven into our daily workflow. Right now the onboarding process feels heavy, and simplifying it would be best.* [Physician, P03.]

In addition to streamlining the clinic workflow, 4 key components were identified as essential for intervention sustainability: executive leadership involvement, integration of operating procedures, personnel, and supplies for patient home testing and monitoring.

Executive leaders, including health system and clinic directors, play a crucial role in setting the vision and strategic direction for sustainability initiatives. Their buy-in enhances advocacy and support, ensuring the integration of BALANCE into clinics and the establishment of policies to embed the intervention into routine operations. Organizational leaders, including clinic managers, can also engage and empower staff by appointing personnel responsible for overseeing and driving intervention sustainability initiatives. Additionally, leadership administrators can ensure the sustainability of necessary equipment by budgeting and sourcing supplies to provide patients with the materials they need for home testing and monitoring. Many participants voiced unmet needs for glucose meters, test strips, and related home-monitoring supplies. In response to these patient-identified needs, clinics should finance glucose meters, test strips, and other testing supplies at the clinic level when patients' insurance does not provide coverage, particularly as DSMES becomes integrated into primary care clinic services. Future implementation addressing these 4 components can contribute to an infrastructure that supports sustainable practices and ensures that incorporating BALANCE is not just an add-on but an integral part of the organization's practice and operations.

Preliminary Efficacy

Patient outcomes on HbA_{1c} levels did not show a statistically significant change (P=.63). Similarly, triglyceride levels decreased from 156.2 mg/dL to 138.1 mg/dL, though not statistically significantly (P=.17). The LDL levels statistically significantly decreased from 77.2 mg/dL to 68 mg/dL (P=.001), indicating a clinically positive impact. The HDL levels slightly decreased from 49.9 mg/dL to 47.1 mg/dL, but this was not statistically significant (P=.15). Diabetes self-care behaviors measured by the SDSCA improved by over 1 point from baseline to 12 months, with exercise performance showing statistical significance (P=.04). Overall, the intervention significantly improved LDL levels and showed positive trends in triglyceride levels and self-care behaviors, suggesting potential efficacy. Refer to Table 3 for detailed outcomes of estimated 12-month differences and Table 4 for detailed analysis of specific self-care behaviors.

XSL•FO
RenderX

Table . Changes in clinical laboratory tests and self-care behavior outcomes.

Health measure	Baseline, mean (SD)	3 months, mean (SD)	6 months, mean (SD)	9 months, mean (SD)	12 months, mean (SD)	Change (baseline to 12 months; 95% CI)	*P* value
HbA_{1c}[a]	7.1 (1.4)	7.1 (1.4)	7.2 (1.1)	7.2 (1.1)	7.2 (0.8)	0.01 (−0.54 to 0.56)	.63
Triglyceride	156.2 (81.9)	139.7 (66.3)	177 (116.6)	157.2 (71.2)	138.1 (69.4)	−22.79 (−55.51 to 9.93)	.17
LDL[b]	77.2 (29.2)	76.4 (29.5)	84.5 (36.7)	69.5 (26.7)	68 (25)	−19.85 (−34.43 to −5.27)	.001
HDL[c]	49.9 (11.6)	50 (12.2)	49.7 (17.9)	47.7 (10.6)	47.1 (12.7)	−3.9 (−8.2 to 0.4)	.15
SDSCA[d]	4.4 (1.5)	5.0 (1.3)	5.0 (1.1)	5.1 (1.2)	5.5 (1.0)	0.38 (−0.58 to 1.34)	.40

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]LDL: low-density lipoprotein.

[c]HDL: high-density lipoprotein.

[d]Diabetes Self-Care Behaviors measured by the Summary of Diabetes Self-Care Activities (SDSCA). Score ranges 0 - 7 with higher score indicating better adherence to diabetes self-care behaviors.

Table . Detailed analysis of specific self-care behaviors of the SDSCA[a].

Diabetes self-care activities	*F* test (*df*)	*P* value
Diet	0.74 (4)	.41
Exercise	5.18 (4)	.04
Glucose monitor	3.97 (4)	.07
Foot care	2.07 (4)	.18
Smoking	1.00 (4)	.34

[a]SDSCA: Summary of Diabetes Self-Care Activities.

The overarching mixed methods domain from integrating the quantitative and qualitative data is reach, efficacy, adoption at the patient level, potential factors related to adoption, and organizational infrastructure for sustainability. Quantitative findings aligned with qualitative results, indicating strong support of BALANCE from both providers and patients, who believed it to be beneficial for the VNAs. The meta-inferences at the patient level are presented in Table 5 and at the clinic level are presented in Table 6. The survey results reflect the entire study sample, whereas the quotations are drawn from a subset of participants, specifically patients who took part in interviews and providers who joined focus groups.

XSL•FO
RenderX

Table . Joint display results of RE-AIM[a] implementation outcomes at the patient level.

Quantitative data summary	Qualitative data summary	Exemplar quote	Meta-inference and interpretations
Reach			
• 10/10 (100%) of eligible clinics participated • 88/175 (50%) of eligible patients participated • Age: 34 - 86 years (mean 67)	The sample had a higher percentage of people who were older (≥65 years) compared to the general population in the geographic area. This is the first diabetes education program offered in the Vietnamese language that leverages mHealth[b] technology and thus it was well received.	• "This program is helpful for old people like me who can't read or understand English." (Patient, 72 years) • "Most Vietnamese patients declined referrals to a DSMES program offered in English, or they came to one session and never came back." (Physician, 57 years)	Most patients were willing to participate in BALANCE[c] because it is available in the Vietnamese language and coincides with their primary care visit rather than coming to separate in-person health education visits.
Efficacy			
• Outcomes: changes from baseline to 12 months, 95% CI (*P* value) • HbA_{1c}[d] (*P*=.63) • Triglyceride (*P*=.17) • LDL[e] (*P*=.001) • HDL[f] (*P*=.15) • SDSCA[g] (*P*=.40)	Patients recognized the importance of changing self-care behaviors. However, many patients had a relapse after the text messaging stopped as some patients felt they were "no longer being watched." Among patients who had improved laboratory test results after completing the intervention, many felt a sense of achievement and then relapsed.	• "I used to think that avoiding sweet food was enough to control my sugar, but I know better now." (Patient, 67 years) • "I eat less rice, and I take my medicines daily now. My blood sugar in the morning has been more stable since I paid attention to what I ate for dinner the night before." (Patient, 56 years)	BALANCE can potentially enhance diabetes-related knowledge, promote health behavior changes, and improve glycemic control. While HbA_{1c} changes were not statistically significant, there was a significant reduction in LDL and improvement in exercise performance.
Adoption at the patient level			
• 79/88 (89.7%) of patients adopted the intervention while 80/88 (90.9%) completed the study (adoption was defined as actively engaged in at least 2 of 3 intervention components).	Facilitators to the intervention adoption at the patient level include the perceived helpfulness and comprehensibility, appropriate mode of delivery, and ease of use. Main barriers to the intervention adoption were the large consumption of phone data for videos and the inability to adopt a new lifestyle because of family responsibilities.	• "The videos are helpful; the messages are nice and short and easy to read." (Patient, 61 years) • "I skipped a few days with videos because it was using up too much of my data." (Patient, 57 years) • "Knowledge is good, but I can't think about diabetes every day; I have to work and take care of my grandchildren." (Patient, 67 years)	Although some patients encountered challenges with viewing text messages/videos, reading printed workbook and/or following the recommended self-care behavior change, there were positive perceptions and experiences, perceived benefits, and willingness to continue with the intervention.

[a]RE-AIM: Reach, Efficacy, Adoption, Implementation, Maintenance.

[b]mHealth: mobile health.

[c]BALANCE: Blended Automated Links Augmented by Nurse Call and Engagement.

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]LDL: low-density lipoprotein.

[f]HDL: high-density lipoprotein.

[g]SDSCA: Summary of Diabetes Self-Care Activities.

XSL•FO
RenderX

Table . Joint display results of PRISM[a] implementation outcomes at the clinic level.

Quantitative data summary	Qualitative data summary	Exemplar quote	Meta-inference and interpretations
Potential factors related to BALANCE[b] adoption			
• Clinics that are independently owned were quicker in initiating the intervention • Clinics that operate within a large health system had a delay in initiating the intervention because of multiple layers in the approval process	All clinic staff and clinicians perceived the intervention as beneficial as long as the clinic staff's workload in implementing the intervention is minimal. Complex referral process and lengthy program enrollment process were perceived as a burden for clinic staff.	• "This program is crucial given the unique barriers that Vietnamese patients face with self-management, but it needs to fit into our workflow." (Physician) • "Transportation have been an issue for our patients with the in-person classes, so text messaging is a great strategy." (Physician)	The need for culturally and linguistically appropriate DSMES[c] was clearly recognized by the providers. However, approvals from health system administrators can influence intervention support and referrals are based on the simplicity of program integration into existing workflow.
Organizational infrastructure for BALANCE sustainability			
Four components were identified as necessary for sustainment: • Buy-in from executive leadership • Program integration with policies and operating procedures • Personnel • Supplies (blood pressure monitors, glucometers, and strips)	The enrollment process takes approximately 30 minutes per patient: complete the consent, demographics, and baseline data. Maintaining this intervention will require either a member of the research team to be present at the clinic or to embed the task in the job description of clinic staff.	• "Implementing it doesn't take much time and effort from our clinic staff once patients are enrolled." (Physician) • "It would be ideal for our EPIC system to integrate with your system to trigger referrals for the program and you take it from there." (Physician)	Providers and clinic staff suggested dedicated personnel, a dedicated day of the week for intervention enrollment, and electronic health records to be interfaced. Another suggestion is to approach large health systems starting at the executive level administrators.

[a]PRISM: Practical, Robust Implementation and Sustainability Model.

[b]BALANCE: Blended Automated Links Augmented by Nurse Call and Engagement.

[c]DSMES: Diabetes Self-Management Education and Support.

Discussion

Principal Findings

The findings from this pilot study indicate that the multicomponent and culturally tailored BALANCE intervention is feasible and acceptable among VNA participants. Out of 175 eligible patients invited to participate in the study, 88 consented, resulting in a 50% enrollment rate. Although this is a lower rate compared to the general population, it is considered to be acceptable, as participation rates in clinical trials among Asian Americans are notably lower compared to other racial and ethnic groups. Studies have shown that Asian Americans are less willing to participate in health research than African Americans, Hispanics or Latinos, and Caucasians [27]. For instance, Asian Americans' participation in clinical trials ranges from 0.75% in cardiovascular trials to 4% in dermatologic trials [28]. This is significantly lower than the participation rates of other groups, such as African Americans, who can account for up to 45% in psychiatric trials [28]. A 50% participation rate and a 90% adoption rate support its feasibility and acceptability. Patient interviews further validated the cultural relevance of the educational materials, which enhanced engagement and contributed to a 100% study completion rate.

Both clinicians and patients reported multiple advantages of BALANCE compared to traditional in-person classes, including eliminating transportation needs, the integration of cultural content, bilingual support, and easy-to-navigate messages. While most patients preferred and appreciated the short daily messages, some reported that the 3-month daily texts with multiple reminders contributed to message fatigue. These findings align with other mHealth intervention research suggesting that simplicity and brevity should be key features of phone-based text messaging design [29,30]. The decline in engagement with mHealth messaging over time is a common trend in digital health interventions. Reasons include message fatigue, reduced perceived relevance, and lack of personalization, which can cause participants to lose interest [31].

Conversely, many patients preferred ongoing messages beyond the 3-month period to maintain engagement and motivation for self-care behavior change, while clinicians suggested expanding BALANCE to support self-management of other chronic conditions, such as hypertension and hyperlipidemia. This underscores the importance of considering individual patient circumstances, such as data availability, when designing mHealth messages. As initial enthusiasm fades, users may deprioritize the messages, especially if they feel repetitive or disconnected from their current needs. Future refinement of BALANCE will include personalizing content, varying message formats, and adding interactive or gamified elements to help sustain interest and encourage continued participation [31,32]. In future work related to this study, a linear mixed-effects model will be used to examine the association between outcome variables and the level of adoption, categorized as high (3 or more engagements) versus low (2 or fewer engagements).

Comparison With Prior Work

In contrast to previous studies where culturally adapted DSMES interventions often showed inconsistent effects on physical activity and participant engagement, these findings highlight

XSL•FO
RenderX

statistically significant improvements in exercise behavior and self-management perceptions among BALANCE participants. Although the overall SDSCA score did not reach significance (P=.40), the intervention improved exercise performance, which showed a statistically significant change over time (P=.04). Furthermore, data from in-depth interviews and focus groups revealed satisfaction with BALANCE where participants regarded it as a positive change in diabetes self-management behaviors. These findings suggest that culturally tailored DSMES interventions can effectively improve self-care behavior and clinical outcomes among study participants but will need additional nudging to maintain desired outcomes. Future refinements will also need to be focused on connecting the relationship between intentions, attitudes, and perceived behavior control.

Our mixed methods analysis revealed 2 significant insights. First, the study team observed a strong connection between culturally tailored DSMES and its acceptance by both clinicians and patients, highlighting how culturally relevant content can enhance engagement and adherence to self-management recommendations. Second, several barriers were identified that are related to the sustainability of BALANCE, including restricted phone data plans and limited resources such as clinic personnel and glucose testing supplies. The need for ongoing commitment and support posed suitable opportunities to maintain the intervention's adoption and sustainment.

Nonadherence to treatment regimen and self-modification of prescribed diabetes medications emerged in in-depth interviews. Specifically, some patients reported reducing their dosage to half of the prescribed amount upon perceiving their blood glucose levels to be within a "normal" range. Cultural beliefs may explain this, as few patients (5%) reported ever receiving DSMES. Evidence indicates that nonadherence to prescribed treatments may lead to poorer diabetes control and increased risk of hospitalization and mortality [33,34]. As a result, diabetes illiteracy exacerbates health disparities among many VNA patients with limited resources for effective self-management. Although BALANCE eliminated the language barriers and improved DSMES access by leveraging mHealth technology, the study team recognized that correcting certain cultural beliefs will be a gradual process.

The HbA_{1c} change was not statistically significant at the 3-month period following the intervention and was slightly increased at the 12-month measurement. It should be noted that the baseline mean HbA_{1c} was 7.1 (SD 1.4), which is very close to the 2025 Standards of Care in Diabetes recommends an A_{1c} goal of less than 7% for most nonpregnant adults [35]. It will be important to assess the efficacy of this intervention in VNAs with poor glycemic control or those with HbA_{1c} higher than 7.5%. Furthermore, the 2025 Standards of Care in Diabetes recommends an LDL cholesterol goal of less than 70 mg/dL for people with diabetes aged 40 to 75 years with higher cardiovascular risk [36]. Participants in this study achieved a mean LDL within the recommended goal range at the 9-month measurement, which was sustained in the 12-month measurement. Lowering LDL levels can reduce diabetes complications such as heart disease, stroke, and kidney disease.

Strengths and Limitations

This study has several strengths, including a mixed methods design integrating programmatic data with in-depth interviews, the first culturally tailored mHealth DSMES intervention for VNAs, and the use of a potentially scalable and accessible mHealth delivery approach. However, this study has some limitations. First, the small sample size makes it challenging to detect statistically significant differences between pre- and postintervention, and a single geographic location limits the generalizability of the findings. However, given the pilot and feasibility nature of this study, these issues are justifiable. Indeed, we demonstrated that it was highly feasible to obtain quantitative RE-AIM and clinical outcomes in this study, which warrants a future fully-powered randomized controlled trial (RCT) to evaluate the BALANCE's efficacy and its implementation. Second, most participants were first-generation immigrants and aged 65 years or older, so the applicability of BALANCE to second-generation immigrants or younger individuals remains to be verified. Third, we did not investigate which education component was most effective at improving outcomes. Future RCTs with a more complex, multiple-assignment design may be needed to address this issue.

Conclusions

The pilot study demonstrated the feasibility, acceptability, and preliminary efficacy of a culturally tailored 3-month intervention using mHealth technologies. Qualitative data complemented the quantitative results, showing that clinic members and patients were overwhelmingly supportive of BALANCE and perceived many benefits. Future research should rigorously evaluate the BALANCE intervention's effectiveness and implementation outcomes through a fully powered RCT. Furthermore, this study supports the potential for wider implementation and scaling of this intervention to improve DSMES access among VNAs and explores long-term sustainability and scalability in different population groups who face similar language and cultural barriers.

Acknowledgments

We extend our sincere gratitude to the dedicated health care providers and clinic staff who generously contributed their time and expertise to this study. Their active participation and insightful feedback were instrumental in refining the intervention, and their commitment to improving patient care is deeply appreciated.

We are grateful to our graduate research assistant, Jimmy Le, for his invaluable contributions to participant recruitment, data collection, and data entry. His efforts were also instrumental in establishing and organizing the study's data collection database.

XSL•FO
RenderX

No generative artificial intelligence (AI) tools were used at any stage in the preparation of this manuscript.

Funding

This research was supported by funding from the Presbyterian Health Foundation through the New Investigator Seed Grant, as well as the Harold Hamm Diabetes Center via the Novel Pilot Grant. TCB is also supported in part by the US NCI Cancer Center Support Grant (P30CA225520) awarded to the University of Oklahoma Stephenson Cancer Center (SCC), and a grant from the Oklahoma Tobacco Settlement Endowment Trust (STCST00400_FY25). We are grateful for their support in advancing innovative approaches to healthcare research. The content is solely the responsibilities of the authors and does not necessarily represent the official views of the funding agencies.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Methods for the National Diabetes Statistics Report. Centers for Disease Control and Prevention. 2023. URL: https://archive.cdc.gov/#/details?url=https://www.cdc.gov/diabetes/php/data-research/methods.html [accessed 2025-06-12]
2. Parker ED, Lin J, Mahoney T, et al. Economic costs of diabetes in the U.S. in 2022. Diabetes Care 2024 Jan 1;47(1):26-43. [doi: 10.2337/dci23-0085] [Medline: 37909353]
3. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943. [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
4. Powers MA, Bardsley JK, Cypress M, et al. Diabetes self-management education and support in adults with type 2 diabetes: a consensus report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. J Am Pharm Assoc (2003) 2020;60(6):e1-e18. [doi: 10.1016/j.japh.2020.04.018] [Medline: 32527704]
5. Davis J, Fischl AH, Beck J, et al. 2022 National Standards for Diabetes Self-Management Education and Support. Sci Diabetes Self Manag Care 2022 Feb;48(1):44-59. [doi: 10.1177/26350106211072203] [Medline: 35049403]
6. Tanaka R, Shibayama T, Sugimoto K, Hidaka K. Diabetes self-management education and support for adults with newly diagnosed type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. Diabetes Res Clin Pract 2020 Nov;169:108480. [doi: 10.1016/j.diabres.2020.108480] [Medline: 33002545]
7. Heron M. Deaths: leading causes for 2018. Natl Vital Stat Rep 2021 May;70(4):1-115. [doi: 10.15620/cdc:104186] [Medline: 34029179]
8. Nguyen A, Edwards K. Barriers and facilitators of diabetes self-management: a qualitative study among Vietnamese Americans. Online J Cult Competence Nurs Healthc 2014;4(2):5-16. [doi: 10.9730/ojccnh.org/v4n2a1]
9. Truong T, Britton M, Harrison D, et al. Assessing the need for diabetes self-management education in the Oklahoma city Vietnamese community. Diabetes Ther 2011 May;2(2):81-91. [doi: 10.1007/s13300-010-0020-6] [Medline: 22127802]
10. De Souza LR, Chan KT, Kobayashi K, Karasiuk A, Fuller-Thomson E. The prevalence and management of diabetes among Vietnamese Americans: a population-based survey of an understudied ethnic group. Chronic Illn 2022 Jun;18(2):306-319. [doi: 10.1177/1742395320959422] [Medline: 33054356]
11. Facts and figures 2020. : International Telecommunication Union; 2020 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf [accessed 2025-07-01]
12. Carroll AE, DiMeglio LA, Stein S, Marrero DG. Using a cell phone-based glucose monitoring system for adolescent diabetes management. Diabetes Educ 2011;37(1):59-66. [doi: 10.1177/0145721710387163] [Medline: 21106908]
13. Franklin VL, Greene SA, Waller A, Greene SA, Pagliari C. Patients' engagement with "Sweet Talk" - a text messaging support system for young people with diabetes. J Med Internet Res 2008 Jun 30;10(2):e20. [doi: 10.2196/jmir.962] [Medline: 18653444]
14. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
15. Mulvaney SA, Anders S, Smith AK, Pittel EJ, Johnson KB. A pilot test of a tailored mobile and web-based diabetes messaging system for adolescents. J Telemed Telecare 2012 Mar;18(2):115-118. [doi: 10.1258/jtt.2011.111006] [Medline: 22383802]

XSL•FO
RenderX

16. Nguyen A, Eschiti V, Bui TC, Nagykaldi Z, Dwyer K. Mobile health interventions to improve health behaviors and healthcare services among Vietnamese individuals: a systematic review. Healthcare (Basel) 2023 Apr 25;11(9):1225. [doi: 10.3390/healthcare11091225] [Medline: 37174767]
17. Ivankova NV, Creswell JW, Stick SL. Using mixed-methods sequential explanatory design: from theory to practice. Field Methods 2006 Feb;18(1):3-20. [doi: 10.1177/1525822X05282260]
18. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf 2008 Apr;34(4):228-243. [doi: 10.1016/s1553-7250(08)34030-6] [Medline: 18468362]
19. Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? Patient Educ Couns 2001 Aug;44(2):119-127. [doi: 10.1016/s0738-3991(00)00186-5] [Medline: 11479052]
20. Belza B, Toobert D, Glasgow R. RE-AIM for Program Planning and Evaluation: Overview and Applications: National Council on Aging; 2007.
21. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
22. Glasgow RE, Dickinson P, Fisher L, et al. Use of RE-AIM to develop a multi-media facilitation tool for the patient-centered medical home. Implement Sci 2011 Oct 21;6(6):118. [doi: 10.1186/1748-5908-6-118] [Medline: 22017791]
23. Fort MP, Manson SM, Glasgow RE. Applying an equity lens to assess context and implementation in public health and health services research and practice using the PRISM framework. Front Health Serv 2023;3:1139788. [doi: 10.3389/frhs.2023.1139788] [Medline: 37125222]
24. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
25. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
26. Assarroudi A, Heshmati Nabavi F, Armat MR, Ebadi A, Vaismoradi M. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. J Res Nurs 2018 Feb;23(1):42-55. [doi: 10.1177/1744987117741667] [Medline: 34394406]
27. Liu Y, Elliott A, Strelnick H, Aguilar-Gaxiola S, Cottler LB. Asian Americans are less willing than other racial groups to participate in health research. J Clin Transl Sci 2019 Jun;3(2-3):90-96. [doi: 10.1017/cts.2019.372] [Medline: 31660231]
28. Lolic M, Araojo R, Okeke M, Temple R. U.S. racial and ethnic participation in global clinical trials by therapeutic areas. J Clin Pharm Ther 2021 Dec;46(6):1576-1581. [doi: 10.1111/jcpt.13532] [Medline: 34544200]
29. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135. [doi: 10.2196/jmir.5066] [Medline: 27287668]
30. So J, Kim S, Cohen H. Message fatigue: conceptual definition, operationalization, and correlates. Commun Monogr 2017 Jan 2;84(1):5-29. [doi: 10.1080/03637751.2016.1250429]
31. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
32. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
33. Ho PM, Rumsfeld JS, Masoudi FA, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-1841. [doi: 10.1001/archinte.166.17.1836] [Medline: 17000939]
34. Lawerence DB, Ragucci KR, Long LB, Parris BS, Helfer LA. Relationship of oral antihyperglycemic (sulfonylurea or metformin) medication adherence aand hemoglobin A1c goal attainment for HMO patients enrolled in a diabetes disease management program. J Manag Care Pharm 2006 Jul;12(6):466-471. [doi: 10.18553/jmcp.2006.12.6.466]
35. ElSayed NA, McCoy RG, Aleppo G, et al. 6. Glycemic goals and hypoglycemia: standards of care in diabetes—2025. Diabetes Care 2025 Jan 1;48(Supplement_1):S128-S145. [doi: 10.2337/dc25-S006]
36. ElSayed NA, McCoy RG, Aleppo G, et al. 10. Cardiovascular disease and risk management: standards of care in diabetes—2025. Diabetes Care 2025 Jan 1;48(Supplement_1):S207-S238. [doi: 10.2337/dc25-S010]

Abbreviations

BALANCE: Blended Automated Links Augmented by Nurse Call and Engagement
DSMES: Diabetes Self-Management Education and Support
HbA_{1c}: hemoglobin A_{1c}
HDL: high-density lipoprotein
LDL: low-density lipoprotein
mHealth: mobile health

XSL•FO
RenderX

PRISM: Practical, Robust Implementation and Sustainability Model
RCT: randomized controlled trial
RE-AIM: Reach, Efficacy, Adoption, Implementation, Maintenance
SDSCA: Summary of Diabetes Self-Care Activities
T2D: type 2 diabetes
VNA: Vietnamese American

Edited by R Correa; submitted 05.Jul.2025; peer-reviewed by E Jones, I Milani, JTS Low; revised version received 21.Feb.2026; accepted 26.Mar.2026; published 07.May.2026.

Please cite as:
Nguyen A, Eschiti V, Bui TC, O'Neal K, Nguyen T, Nguyen AP, Li H, Machiorlatti M, Dwyer K
Promoting Diabetes Self-Management Among Vietnamese Americans: Mixed Methods Pilot Study
JMIR Diabetes 2026;11:e80177
URL: https://diabetes.jmir.org/2026/1/e80177
doi:10.2196/80177

© Anna Nguyen, Valerie Eschiti, Thanh C Bui, Katherine O'Neal, Tam Nguyen, Angelina P Nguyen, Hanxia Li, Michael Machiorlatti, Kathleen Dwyer. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 7.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Promoting Stigma-Free Language in Type 2 Diabetes Care Among Primary Care Clinicians: Randomized Controlled Pilot Trial

Kevin L Joiner[1], PhD; Alexandra Agapiou[2], DNP; Alicia Carmichael[3], MS; Brandon P Labbree[3], MA; Donna Walter[3], BS; Houda Tarraf[3], MSW; Medha Raju[3], MA; Boluwatife Dogari[3], BA; Nao Hagiwara[4], PhD; Sophia Lin[1]; Richard Gonzalez[3], PhD

[1]Department of Health Behavior and Clinical Sciences, School of Nursing, University of Michigan, 426 N Ingalls St, Ann Arbor, MI, United States

[2]Department of Systems, Populations, and Leadership, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[3]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States

[4]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Kevin L Joiner, PhD
Department of Health Behavior and Clinical Sciences, School of Nursing, University of Michigan, 426 N Ingalls St, Ann Arbor, MI, United States

Abstract

Background: The American Diabetes Association and the Association of Diabetes Care & Education Specialists recommend person-first, strengths-based, and stigma-free language in diabetes care and education. However, interventions to promote stigma-free language among primary care clinicians remain limited.

Objective: This mixed methods pilot study aimed to evaluate the acceptability and preliminary effects of a diabetes stigma-reduction training module for primary care clinicians, designed to increase awareness of language used in clinician-patient interactions and explore changes in attitudes and intentions toward avoiding stigmatizing language.

Methods: Twenty-one primary care clinicians completed either a stigma-reduction training module (n=12) or an active control condition (n=9). Both conditions included a standardized patient encounter, an educational video, and a guided self-reflection activity. The stigma-reduction training module included diabetes-specific content on stigmatizing language and person-first, strengths-based alternatives. Theory of Planned Behavior domain measures were administered before and immediately after the assigned condition. Attitudes and intentions were assessed precondition and postcondition. Acceptability was evaluated using a postcondition survey and semistructured qualitative interviews. Repeated-measures ANOVA was used to assess between-group changes over time, and a qualitative descriptive approach was used to explore clinicians' perceptions.

Results: Clinicians who received the stigma-reduction training module demonstrated significantly greater improvement in attitudes toward avoiding stigmatizing language compared with clinicians in the active control condition, with mean attitude scores increasing by 1.19 (SD 0.68) points versus 0.21 (SD 0.44) points, respectively. A significant condition × time interaction was observed ($F_{1,19}$=14.27; P=.001, partial η^2=0.43, Cohen d=1.67). The between-group difference in intention change was not statistically significant, with mean intention scores increasing by 0.69 (SD 1.08) points in the stigma-reduction training module condition versus 0.02 (SD 0.90) points in the active control condition ($F_{1,19}$=2.33; P=.14, partial η^2=0.11, Cohen d=0.67). Acceptability ratings were higher for the stigma-reduction training module than for the active control condition for clarity (mean 6.17, SD 0.72 vs mean 4.44, SD 1.74; P=.02), helpfulness (mean 6.08, SD 1.00 vs mean 4.56, SD 1.59; P=.03), and likelihood of recommendation (mean 6.25, SD 0.87 vs mean 4.56, SD 1.67; P=.02). The qualitative findings suggested that clinicians perceived the stigma-reduction training module as actionable, relevant to primary care, and useful for promoting self-reflection about habitual language use. The active control condition was generally perceived as a useful refresher but less novel and impactful.

Conclusions: This pilot study suggests that a brief, theory-informed diabetes stigma-reduction training module is acceptable and associated with improved clinician attitudes toward stigmatizing language in a simulated clinical environment. Changes in intentions were descriptively in the expected direction but did not reach statistical significance. The findings should be interpreted cautiously, given the small sample size and lack of behavioral outcome assessment. Future studies should evaluate the module in larger, more diverse samples and assess whether changes in attitudes translate into sustained changes in clinician language use and patient experiences.

(JMIR Diabetes 2026;11:e95805) doi:10.2196/95805

XSL•FO
RenderX

KEYWORDS

type 2 diabetes; diabetes stigma; health care communication; clinician attitudes; medical education; intervention; patient-provider interaction

Introduction

Over 38 million US adults are living with diabetes, and approximately 90% to 95% of them have type 2 diabetes (T2D) [1]. Although adults with T2D strive to lead full, productive, and healthy lives, T2D remains a complex and demanding chronic condition. Recent advances in treatments and technologies have helped prevent or delay diabetes-related complications and mortality while improving health and quality of life [2]. However, many US adults with T2D continue to face significant challenges, underscoring the need for ongoing efforts to improve diabetes care, self-management education, and support [3].

Emerging evidence suggests that adults with T2D encounter diabetes-related stigma in social, work, and health care settings [4,5]. Stigma is a social process in which an identifiable characteristic is used to label individuals or groups as distinct from a dominant reference group, typically in the context of unequal social power [6]. For people with T2D, stigma may arise from the diagnosis itself as well as from the pervasive societal judgments, assumptions, and stereotypes associated with the condition [7]. Stigma is commonly conceptualized as social stigma and self-stigma [6]. Social stigma includes perceived or experienced blame, judgment, stereotyping, rejection, exclusion, or discrimination, whereas self-stigma occurs when individuals internalize stigmatizing beliefs, leading to feelings of embarrassment or shame [6].

Among adults with T2D, diabetes-related stigma has been associated with greater diabetes distress, elevated depressive and anxiety symptoms, reduced engagement in self-care, diminished self-efficacy, poorer health care experiences, and reduced health care use [7-9]. Although research on diabetes stigma in health care settings remains limited, these findings suggest that clinician-patient communication may be an important target for stigma-reduction efforts [10].

Minimizing stigmatizing language in clinician-patient communication has therefore emerged as a priority in diabetes care [11]. According to the National Institutes of Health frameworks, clinician-patient communication is a key factor at the intersection of personal and system-level determinants [12]. Because most adults with T2D receive their diabetes care in primary care settings, effective communication between primary care clinicians and adults with T2D is essential for supporting self-management, addressing logistical and structural barriers to care, and facilitating appropriate initiation and intensification of therapies [13-16].

In 2017, the American Diabetes Association (ADA) and the American Association of Diabetes Educators, now the Association of Diabetes Care & Education Specialists (ADCES), issued a joint position statement on the use of language in diabetes care, recommending person-first, strengths-based, and stigma-free communication. This guidance is foundational to reducing diabetes stigma in clinical care [13,17]. However, the original statement appeared primarily in specialty journals, potentially limiting awareness among primary care clinicians. Primary care clinicians are therefore an important target for intervention because they play a central role in diabetes care and may be well positioned to shift clinician-patient communication away from language that contributes to diabetes stigma [15]. Nevertheless, many primary care clinicians may be insufficiently aware of the detrimental effects of diabetes stigma or the ways in which routine clinical language can perpetuate harm [11].

To address this gap, we developed a diabetes stigma-reduction training module designed to raise awareness of the ADA and ADCES language guidance and promote self-reflection on language used in primary care encounters. The module includes 3 sequential activities: a standardized patient encounter to capture language used in a simulated clinical interaction, an educational video providing guidance and examples of stigmatizing language and stigma-free alternatives, and a guided self-reflection activity in which clinicians review their recorded encounter and identify opportunities for language improvement. The module was grounded in the Theory of Planned Behavior, which posits that intentions to perform behaviors, such as avoiding stigmatizing language, are shaped by attitudes, subjective norms, and perceived behavioral control [18]. The module was also informed by growing recognition that stigma reduction efforts should look beyond a single disease or condition and address mechanisms that operate across multiple conditions [19,20]. Consistent with prior research in stigma reduction in health care, the module combined informational content with skills-building through role-play and structured, nonjudgmental reflection on personal communication behaviors [19,21].

The objective of this pilot study was to evaluate the acceptability of the diabetes stigma-reduction training module among primary care clinicians and to explore preliminary changes in attitudes and intentions toward avoiding stigmatizing language in clinician-patient interactions.

Methods

Study Design

This randomized controlled pilot trial was conducted and reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials; Checklist 1) 2025 statement[22]. The pilot study was a randomized controlled trial using a convergent mixed methods design. Quantitative surveys and qualitative exit interviews were used to evaluate the acceptability and preliminary effects of the stigma-reduction training activities for primary care clinicians and to inform the design of a larger, multisite trial.

Participants and Recruitment

Because all study activities were conducted in person, recruitment focused on the local geographic area. Targeted

XSL•FO
RenderX

emails with informational flyers were sent to potentially eligible individuals listed in a registry of persons interested in participating in research studies maintained by the Michigan Institute for Clinical & Health Research [23]. Letters with informational flyers were also mailed to primary care clinicians practicing in the local region. In addition, flyers were posted on professional listservs and digital displays in local health systems.

Eligible participants were practicing clinicians, including physicians, nurse practitioners, and physician assistants, who provided outpatient primary care for adults with T2D on a regular basis. Eligibility was confirmed through an online screening survey.

Procedures and Setting

All study activities took place in a simulated clinical environment at the University of Michigan HomeLab simulation facility (RRID:CR_026753), a 985-square-foot space designed to be modified into realistic living and working environments [24]. For this study, the facility was configured to mimic an outpatient clinical setting and create an immersive experience for participating primary care clinicians. Participant-facing activities occurred in a private survey and interview room, a simulated clinic hallway, and a simulated patient examination room. An adjacent audio-visual control room was used by study staff to operate recording equipment. To enhance realism, a simulated paper-based medical chart was used containing fabricated laboratory results, progress notes, and patient questionnaires.

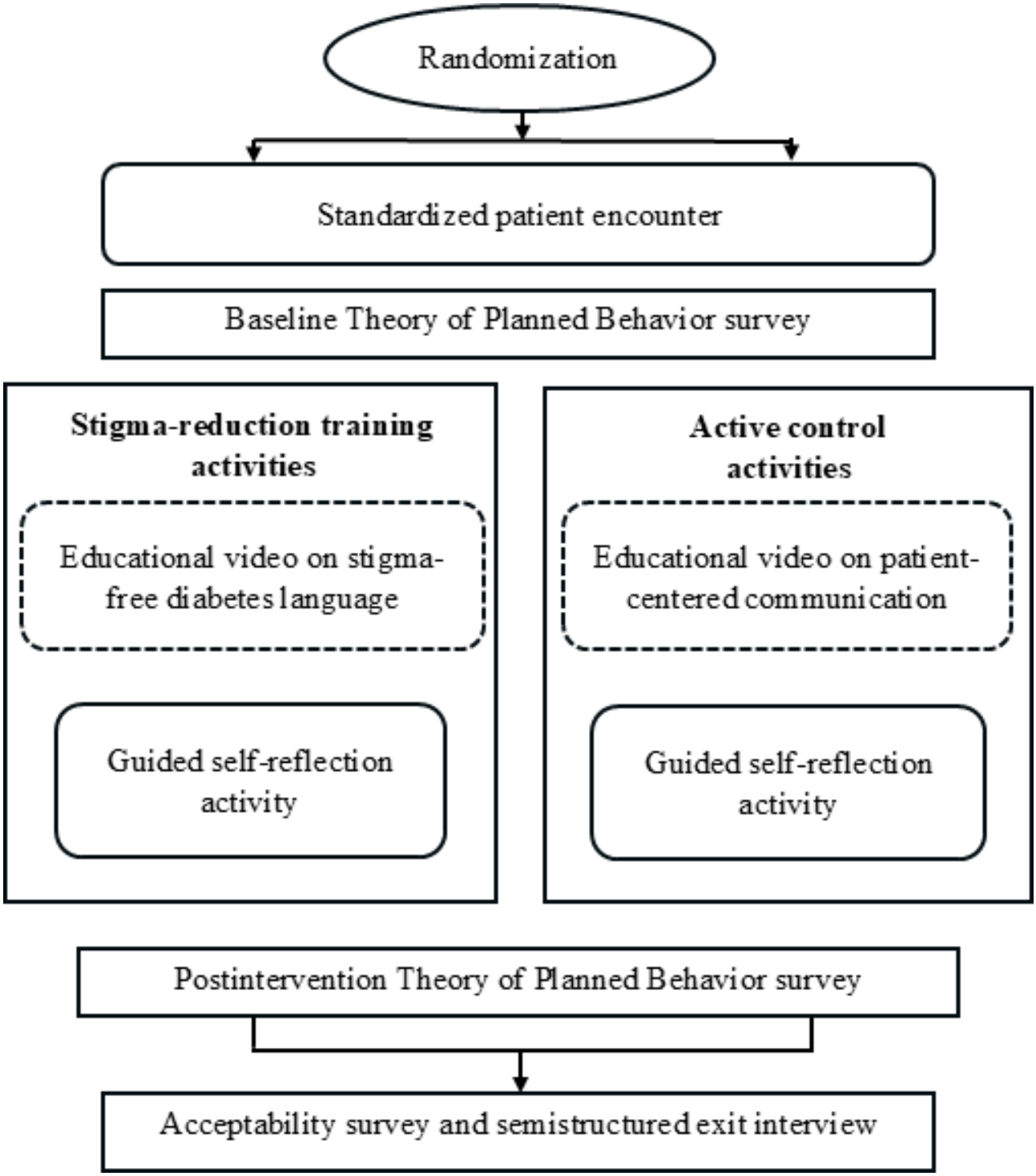
A computer-aided lab interviewing (CALI) system, programmed within Qualtrics, was used to promote protocol standardization and maintain allocation concealment. The CALI was operated on a computer tablet by a single study team member, who was the only member of the study team to have direct contact with the participant during the in-person visit. The recruitment coordinator had contact with participants before the study, either online or by telephone. Because the randomized condition was preprogrammed by the recruitment coordinator, the research team member with direct participant contact was blinded to the participants' randomized allocation during the study. Quantitative outcomes were collected through participant-completed electronic surveys and scored according to prespecified rules. Therefore, there was no separate human outcome assessor for the quantitative outcomes. Data analysts were not blinded during final statistical modeling because condition labels were required for the planned between-group analyses.

Potential participants completed an online screener. Once eligibility was confirmed by the recruitment coordinator, the participant consented and enrolled in the study. Following consent and enrollment, the recruitment coordinator randomly allocated the participants to the stigma-reduction training module or the active control condition using a computer-generated randomization sequence. The computer-generated randomization sequence was stored by the recruitment coordinator in a secure electronic file. Following randomization, the recruitment coordinator entered the participant's unique ID number and their randomized condition into the CALI.

Participating clinicians arrived in the building lobby, where they were greeted by the research staff member who implemented the study protocol and escorted to the simulation space. This research staff member was the only member of the study staff, other than the recruitment coordinator, with direct participant contact. The CALI was used by the research staff member to implement the study protocol with the participant. The CALI contained standardized scripts, reminders, and checklists to ensure consistent instructions and procedures across participants. The study visit flow is shown in Figure 1, and the full study visit protocol and choreography are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Study visit flow and randomized condition sequence.



Intervention

All participants first completed a standardized patient encounter to capture baseline communication in a simulated clinical interaction. Participants then completed either the stigma-reduction training activities or the active control activities, consisting of an educational video followed by guided self-reflection using the encounter recording and transcript. The key difference between the stigma-reduction training activities and the active control activities was the educational video.

All participants completed a 15-minute role-play consultation in the simulated patient examination room with a trained actor portraying an adult with T2D attending a primary care visit for T2D with a new clinician. Two trained actors, 1 man and 1 woman, portrayed the standardized patient across study visits. Simulated paper medical charts included laboratory results, progress notes, and health questionnaires. Encounters were audio- and video-recorded and transcribed to capture natural patterns of clinician-patient communication.

Participants assigned to the stigma-reduction training module viewed a 13-minute educational video developed by members

XSL•FO
RenderX

of the study team [25]. The video provided an overview of ADA and ADCES language guidance and featured contrasting vignettes illustrating stigmatizing terms and phrases alongside person-first, strengths-based alternatives. Participants assigned to the active control condition viewed a 13-minute publicly available video on general principles of patient-centered communication that did not address diabetes or stigma [26].

In both conditions, participants completed the same guided self-reflection procedure using their own encounter recording and transcript. Participants reviewed their standardized patient encounter in the private survey and interview room and identified opportunities to improve their communication based on the educational video they had viewed. Thus, the focus of reflection differed by condition because participants applied either the diabetes stigma-reduction video content or the general patient-centered communication video content. The guided self-reflection activity script and procedures are provided in Multimedia Appendix 2. In the stigma-reduction condition, this confrontation-inspired approach was intended to engage participants in recognizing their own communication patterns and forming intentions to avoid stigmatizing language in their primary care practice in the future [27-29].

The stigma-reduction training module was grounded in the Theory of Planned Behavior [18]. The educational video was designed to target attitudes by reframing stigmatizing language from routine clinical terminology to language with potential clinical harm. The guided self-reflection activity was designed to promote awareness of habitual communication patterns; in the stigma-reduction condition, this reflection was paired with educational content emphasizing stigma-free diabetes language.

Theory of Planned Behavior Domain Measures

Preliminary effects were evaluated using surveys before and after the assigned educational video and guided self-reflection activities. The baseline survey was administered after the standardized patient encounter and before participants viewed the assigned educational video or completed the guided self-reflection. The postintervention survey was administered after the completion of the educational video and guided self-reflection. The survey was developed for the study to measure Theory of Planned Behavior domains related to stigmatizing language use in clinician-patient interactions during routine primary care for type 2 diabetes [30]. The full instrument and scoring procedures are provided in Multimedia Appendix 3. At baseline, the survey assessed 5 domains: self-reported behavior, attitudes, subjective norms, perceived behavioral control, and intentions. Immediately after exposure to the assigned educational video and guided self-reflection activity, attitudes and intentions were reassessed. These domains were prioritized for postintervention assessment because the brief study window was expected to be most sensitive to changes in attitudes and intentions, whereas subjective norms, perceived behavioral control, and self-reported behavior were not expected to change meaningfully over this period.

The Theory of Planned Behavior survey included 6 language banks: 5 representing nonrecommended or stigmatizing language and 1 representing recommended person-first, strengths-based language based on the ADA and ADCES guidance statement [17]. The 5 nonrecommended language banks included authoritative language, such as “compliant,” “non-compliant,” “compliance,” and “non-compliance”; non-person-centered language, such as “diabetic” used as a noun, for example, “Are you a diabetic?”; morally implicative language, such as “control” used as a verb, adjective, or noun, including “controlled,” “uncontrolled,” “glucose control,” and “glycemic control”; negative connotation language, such as “poor,” “poorly,” “bad,” “badly,” “fail,” and “failed”; and threatening language, such as statements focused on loss, including “You are going to end up on dialysis,” “You are going to end up blind,” or “You are going to lose a foot.” The recommended language bank included strengths-based statements focused on gain, such as “More and more people are living long and healthy lives with diabetes. Let’s work together to make a plan that you can do in your daily life.”

For each language bank, participants responded to 5 Theory of Planned Behavior domains using 7-point Likert scales. One item assessed self-reported behavior: “During the past 4 weeks, about how often have you used these words in your appointments with people with type 2 diabetes?” with anchors ranging from “all the time” to “never.” Two semantic differential items assessed attitudes: “For me, using these words in my appointments with people with type 2 diabetes is,” with anchors “good-bad” and “useful-worthless.” One item assessed subjective norms: “Most health care providers like me use these words in their appointments with people with type 2 diabetes,” with anchors “strongly agree-strongly disagree.” One item assessed perceived behavioral control: “My using or not using these words in my appointments with people with type 2 diabetes, is completely up to me,” with anchors “strongly agree-strongly disagree.” One item assessed intentions: “I intend to use these words in my appointments with people with type 2 diabetes, in the next 6 months,” with anchors “strongly agree-strongly disagree.”

Domain scores were calculated by averaging the relevant items across the 6 language banks. For the self-reported behavior, attitude, subjective norm, and intention domains, the strengths-based language bank was reverse-scored so that higher scores consistently reflected movement away from stigmatizing language and toward person-first, strengths-based language. For perceived behavioral control, all items were reverse-coded so that higher scores indicated greater perceived control over language use.

Acceptability Measures and Qualitative Exit Interview

Acceptability was assessed using quantitative and qualitative approaches. Immediately after completing the standardized patient encounter, assigned educational video, guided self-reflection activity, and postintervention Theory of Planned Behavior survey, clinicians completed a brief 4-item acceptability survey adapted from prior work [31]. Before completing the acceptability survey, all participants were provided with a brief handout summarizing the ADA and ADCES diabetes language recommendations and were instructed to rate the study activities in relation to bringing awareness to those recommendations. Items assessed the perceived amount of information, clarity of activities, helpfulness of activities,

XSL•FO
RenderX

and likelihood of recommending the activities to other clinicians. The acceptability survey is provided in Multimedia Appendix 4. All four items were rated on 7-point scales. For the amount of information item, response options ranged from 1, "too little information," to 4, "just the right amount of information," to 7, "too much information." For clarity, helpfulness, and likelihood of recommendation, higher scores indicated greater acceptability of the stigma-reduction training module or the active control condition.

Semistructured exit interviews were conducted to explore clinicians' prior exposure to ADA and ADCES language guidance, perceived barriers and facilitators to accessing and implementing these recommendations, and feedback on each activity in the stigma-reduction training module or the active control condition. The semistructured interview guide is provided in Multimedia Appendix 5. The interviews were recorded and transcribed for qualitative analysis. Demographic and clinical practice characteristics were collected online before the study visit.

Statistical Analysis

Analyses were conducted using R (version 4.4.0; R Core Team). Analyses are presented according to condition received. Sensitivity analyses comparing as-randomized, as-treated, and per-protocol approaches yielded similar conclusions; therefore, as-treated results are reported. Means and SDs were used to describe continuous variables. Participant demographic and clinical practice characteristics were compared by condition received using Mann-Whitney *U* tests for continuous variables and Fisher exact tests for categorical variables, implemented with the *gtsummary* package.

Changes in Theory of Planned Behavior attitude and intention scores from baseline to postintervention were analyzed using 2 × 2 repeated-measures ANOVAs, with condition as the between-subjects factor and time as the within-subjects factor, implemented with the *afex* package. Planned contrasts were used to compare between-group differences at baseline and within-group pre-post changes, or simple main effects, using the *emmeans* package. Effect sizes were calculated as Cohen *d* and partial η^2 using the *effectsize* package. Between-group differences in acceptability scores were compared using independent-samples *t* tests, with Cohen *d* calculated as the effect-size measure.

Qualitative Analysis

All interviews were audio-recorded and machine-transcribed verbatim. Qualitative analysis followed a qualitative descriptive approach [32]. Two authors (AA and SL) independently reviewed and coded the transcripts, which were labeled only by participant ID. Study condition labels were withheld during coding; however, because transcripts were verbatim, analysts may have inferred condition from participants' comments about the educational content. An initial codebook was developed through discussion with the principal investigator (KLJ) and refined iteratively. All transcripts were coded using ATLAS.ti Web (version 5.8.0). Coding discrepancies were resolved through discussion until consensus was reached. After coding was complete, codes were grouped into preliminary themes. The team refined themes through discussion, clarified theme definitions, and selected representative quotations. Years of primary care practice were aggregated into categories to protect participant confidentiality.

Ethical Considerations

Participants who completed a study visit received an honorarium of US $80-$100, parking reimbursement, and entry into a raffle for a sports memorabilia item valued at US $250. The raffle was conducted after all participants had completed study participation, and the winner was notified by the recruitment coordinator. The protocol was deemed exempt by the University of Michigan Institutional Review Board (HUM00235534).

Results

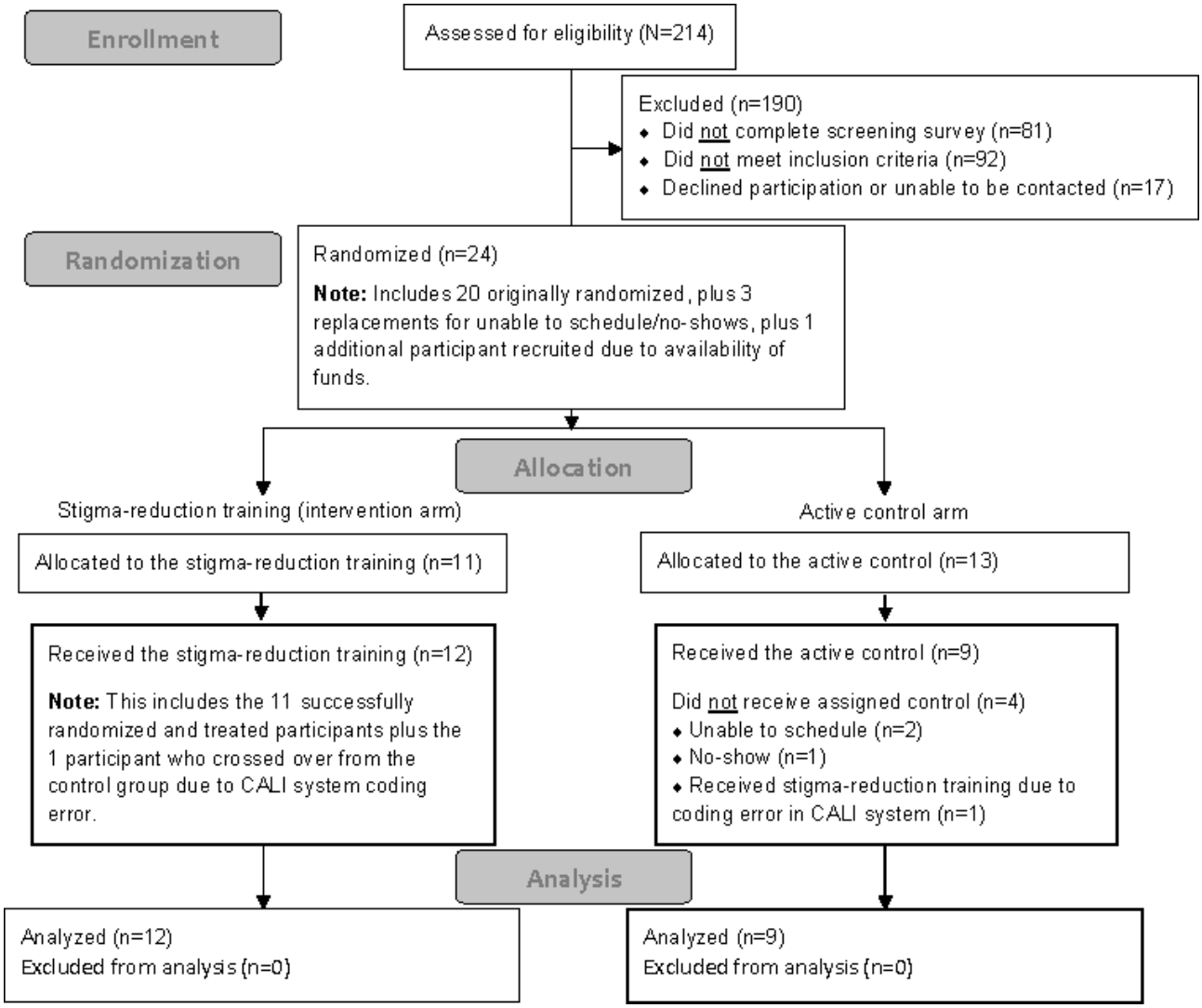
Participant Flow and Characteristics

Clinicians were recruited and participated in the study between July 2023 and July 2024. A total of 214 individuals initiated the screening survey during this period (Figure 2). Of these, 190 were excluded: 81 did not complete the screening survey, 92 did not meet eligibility criteria, and 17 declined participation or could not be contacted. The original study design called for 20 participants randomly allocated 1:1 to the stigma-reduction training module or the active control condition. Because this was an early-stage pilot study, the sample size was selected to estimate acceptability and preliminary outcome patterns rather than to test definitive efficacy, consistent with guidance for pilot studies [33]. Before initiating a study visit, 3 enrolled participants left the study: 2 could not be scheduled, and one did not attend the scheduled visit and was subsequently lost to follow-up. Three replacement participants were recruited and randomized. In addition, sufficient remaining funds allowed the recruitment of one additional participant. Ultimately, 21 clinicians completed a study visit and were included in the analysis.

Due to a programming error in the electronic system used to implement the randomized study visit sequence, one participant randomly allocated to the active control condition received the stigma-reduction training module. Sensitivity analyses were conducted to assess the robustness of the findings to this crossover. In the as-randomized analysis, participants were analyzed according to their original randomized allocation. In the as-treated analysis, participants were analyzed according to the condition received. In the per-protocol analysis, the participant affected by the programming error was excluded. These analyses yielded consistent conclusions for attitudes. For intentions, the as-treated and per-protocol analyses were not statistically significant, whereas the as-randomized analysis reached statistical significance; therefore, intention findings were interpreted cautiously. The results of the sensitivity analyses are provided in Multimedia Appendix 6. Of the 21 clinicians who completed the study, 12 received the stigma-reduction training module, and 9 received the active control condition.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. CALI: computer-aided lab interviewing.



Participant demographic and clinical practice characteristics are summarized in Table 1. The mean age of participants was 38 (SD 12) years, and most participants identified as women. Participants identified predominantly as White, followed by Asian, Middle Eastern or North African, Multiracial, and Hispanic or Latino. On average, clinicians reported 6 patients with T2D per day in clinical practice and had 7 years of clinical experience. Most were physicians, followed by nurse practitioners and physician assistants. There were no statistically significant differences between study conditions for any demographic or clinical practice characteristic.

XSL•FO
RenderX

Table . Demographic and clinical practice characteristics of primary care clinician participants by condition received.

Characteristic	Overall (n=21)	Stigma-reduction training module (n=12)	Active control condition (n=9)	*P* value[a]
Demographic characteristics				
Age (y), mean (SD)	38 (12)	40 (13)	37 (11)	.54
Gender identity, n				>.99
Man	5	3	2	
Woman	16	9	7	
Race or ethnicity, n				.43
Asian	3	3	0	
Hispanic or Latino	1	1	0	
Middle Eastern or North African	3	2	1	
Multiracial	2	1	1	
White	12	5	7	
Clinical practice characteristics				
Adults with T2D[b] seen on a typical day, number, mean (SD)	6 (5)	7 (6)	4 (2)	.37
Practice setting, n				.62
Suburban	11	6	5	
Urban	5	2	3	
Multiple	5	4	1	
Primary care practice (y), mean (SD)	7 (8)	6 (7)	8 (9)	.94
Medical profession, n				.80
Physician	13	8	5	
Nurse practitioner	7	4	3	
Physician assistant	1	0	1	

[a]Mann-Whitney *U* tests were used for continuous variables, and Fisher exact tests were used for categorical variables.

[b]T2D: type 2 diabetes.

Quantitative Findings

Separate 2 × 2 mixed ANOVAs were used to assess changes in attitudes and intention scores regarding stigmatizing language (Table 2). Participants exposed to the stigma-reduction training module demonstrated significantly greater improvement in attitude scores over time than participants in the active control condition, as indicated by a significant condition-by-time interaction ($F_{1,19}$=14.27; *P*=.001). This represented a large effect (Cohen *d*=1.67, partial η^2=0.43). For intention scores, the between-group difference in change over time was not statistically significant ($F_{1,19}$=2.33; *P*=.14). However, the effect was in the expected direction, with greater improvement in the stigma-reduction training module condition. The magnitude of this effect was medium (Cohen *d*=0.67, partial η^2=0.11).

Acceptability ratings were higher among participants who were exposed to the stigma-reduction training module for clarity, helpfulness, and likelihood of recommendation (all *P*<.05; Table 3). Perceptions of the amount of information conveyed did not differ significantly between groups (*P*=.37), with both groups rating the amount of information close to "just the right amount."

XSL•FO
RenderX

Table . Theory of Planned Behavior domain scores by condition received[a].

Domain and condition	Baseline score	*P* value[b]	Postcondition score	Change from baseline	*P* value[c]	Partial η^2[c] (95% CI)	Cohen *d*[c] (95% CI)
Self-reported behavior		.41			N/A[d]	N/A	N/A
Stigma-reduction module, mean (SD)	4.57 (0.64)		__[e]	N/A			
Active control, mean (SD)	4.74 (0.80)		__[e]	N/A			
Attitudes		.06			.001	0.43 (0.15-1.00)	1.67 (0.64 to 2.66)
Stigma-reduction module, mean (SD)	4.61 (0.53)		5.80 (0.68)	1.19 (0.68)			
Active control, mean (SD)	5.28 (0.80)		5.49 (0.87)	0.21 (0.44)			
Subjective norms		.75			N/A	N/A	N/A
Stigma-reduction module, mean (SD)	3.39 (0.72)		__[e]	N/A			
Active control, mean (SD)	3.46 (0.78)		__[e]	N/A			
Perceived behavioral control		.66			N/A	N/A	N/A
Stigma-reduction module, mean (SD)	6.39 (0.80)		__[e]	N/A			
Active control, mean (SD)	6.24 (0.84)		__[e]	N/A			
Intentions		.17			.14	0.11 (0.00-1.00)	0.67 (−0.22 to 1.56)
Stigma-reduction module, mean (SD)	3.69 (0.76)		4.39 (0.98)	0.69 (1.08)			
Active control, mean (SD)	4.17 (0.50)		4.19 (1.05)	0.02 (0.90)			

[a]Attitudes and intentions were assessed after the standardized patient encounter, before the assigned educational video and guided self-reflection activity, and again immediately after those activities. Self-reported behavior, subjective norms, and perceived behavioral control were assessed at baseline only.

[b]Baseline between-group comparison.

[c]The *P* value and partial η^2 are from the condition × time interaction. Cohen *d* represents the between-group difference in change scores.

[d]Not applicable.

[e]Not available.

XSL•FO
RenderX

Table . Acceptability scores by condition received[a].

Acceptability component	Stigma-reduction module (n=12), mean (SD)	Active control (n=9), mean (SD)	*P* value[b]
Amount of information	3.75 (0.62)	3.33 (1.22)	.37
Clarity of activities	6.17 (0.72)	4.44 (1.74)	.02
Helpfulness of activities	6.08 (1.00)	4.56 (1.59)	.03
Likelihood of recommending	6.25 (0.87)	4.56 (1.67)	.02

[a]All acceptability components were rated on 7-point scales.

[b]*P* values are from independent-samples *t* tests.

Qualitative Findings

The semistructured exit interviews highlighted three main themes: (1) Integration of communication strategies in clinical practice; (2) Engagement in self-reflection and pursuit of continuous improvement; and (3) Implementation challenges, convenience, and engagement.

Theme 1: Integration of Communication Strategies in Clinical Practice

Several clinicians in the stigma-reduction training module arm described the module as providing education and opportunities for self-reflection to help them consider avoiding stigmatizing language and using stigma-free alternatives in clinical practice. Some commented on the clear, pragmatic approach of the stigma-reduction training module, noting that the content aligned closely with primary care realities and evolving standards. As one clinician explained:

> *You want to know what you're doing actually has an impact. Sometimes…this kind of thing feels nebulous, but this training felt actionable, like something you could really use.* [Male, physician, primary care practice 5 - 10 years]

Several described the stigma-reduction training module as reflecting a broader transformation in health care norms. One participant stated:

> *It's a paradigm shift…Terms like "compliance" are being phased out for "adherence," and that's really important.* [Male, physician, primary care practice less than 5 years]

Such comments highlight how, for some participants, the intervention aligned with their interest in language choices that reflect changes in practice.

By contrast, some clinicians who received the active control condition saw the content as largely familiar and less likely to influence established routines. One participant summarized this perspective:

> *The video was a reminder of the basics…. At my level, it felt like things I'd seen plenty of times…helpful for students, but not very new.* [Physician, primary care practice less than 5 years]

This suggests that, while general communication refreshers have value, targeted, recommendation-based interventions may do more to catalyze meaningful practice change.

Theme 2: Engagement in Self-Reflection and Pursuit of Continuous Improvement

Some participants who received the stigma-reduction training module commented on the value of self-reflection promoted by reviewing their own recorded encounters. Several described increased awareness of their habitual language and a willingness to alter established communication patterns. One participant stated:

> *Looking at my own speech, I realized how much I use words like "uncontrolled."…I often rely on scare tactics, thinking they'll motivate patients.…This made me see that being more positive is much more impactful.* [Female, physician, primary care practice more than 10 years]

One participant expressed that the stigma-reduction training module fostered immediate motivation for change:

> *Using neutral language…instead of negative warnings creates more productive conversations.…I can start changing that tomorrow.* [Female, physician, primary care practice less than 5 years]

One clinician with extensive practice experience in primary care described the activities in the stigma-reduction training module as providing "more meaningful insights and opportunities for reflection" than didactic training alone (female, physician, primary care practice more than 10 years). This suggests that experiential learning, such as video and transcript review, has the potential to promote self-awareness.

Among those who received the active control condition, some participants said that even when they recognized the usefulness of improved communication, practical limitations often hampered their ability to change. One clinician expressed:

> *Usually, I'll say, "How does this plan sound to you?"…But visits are so rushed…15 minutes each, back-to-back…it's hard to be as thorough as these recommendations suggest.* [Female, physician, primary care practice less than 5 years]

Comments such as this underscore the role of systemic and contextual challenges.

Theme 3: Implementation Challenges, Convenience, and Engagement

Some participants reflected on both logistical and motivational factors affecting engagement in continuing education and professional development. Several clinicians who received the

XSL•FO
RenderX

stigma-reduction training module commented on the value of structured opportunities to improve communication style. One stated:

> *The farther you get from training, the fewer chances you have to step back and assess yourself.…You get stuck in your routines and stop thinking about your delivery.* [Female, physician, primary care practice 5 - 10 years]

Several highlighted the importance of authoritative guidance in motivating real change:

> *Having the actual guidelines helped.…With just opinions, I'm skeptical, but guidelines make me more willing to change my language.* [Female, physician, primary care practice less than 5 years]

Conversely, one participant who received the control condition questioned the added benefit of the general communication training module amid competing priorities.

> *Honestly, at this stage, it's mostly reinforcing things I already do.…It might work for students, but for most of us the issue really comes down to lack of time.* [Male, physician, primary care practice less than 5 years]

Even so, one of the clinicians who received the control condition agreed that reviewing their own standardized patient encounter was valuable.

> *Seeing my body language and eye contact from the patient's point of view…honestly, that was the most helpful part.* [Female, nurse practitioner, primary care practice 5 - 10 years]

Discussion

Principal Findings

This pilot study suggests that a brief, theory-informed stigma-reduction training module is acceptable and may improve primary care clinicians' attitudes toward avoiding stigmatizing language in T2D care. Compared with the active control condition, the stigma-reduction training module was rated significantly higher in clarity, helpfulness, and likelihood of recommendation. The survey findings showed a significant improvement in attitudes toward avoiding stigmatizing language among clinicians who received the stigma-reduction training module. Changes in intentions were descriptively in the expected direction but did not reach statistical significance.

The primary quantitative finding, a significantly greater improvement in attitudes in the stigma-reduction training module condition compared with the active control condition, aligns with the Theory of Planned Behavior. Within this framework, attitudes are a key antecedent of behavioral intention. The observed improvement suggests that the module may have shifted how clinicians appraise stigmatizing language, from routine or familiar clinical terminology to language with potential to cause harm. Although the between-group change in intentions did not reach statistical significance, the descriptive pattern was in the expected direction. This finding should be interpreted cautiously and does not provide evidence that the module changed clinicians' intentions.

The qualitative findings helped to explain these quantitative results. Clinicians who received the stigma-reduction training module described the training as practical, relevant, and immediately applicable to clinical care. They emphasized the value of concrete language alternatives and noted that the module aligned with evolving professional standards for person-centered diabetes communication. Several participants described the guided self-reflection activity as especially useful because reviewing their own recorded standardized patient encounters allowed them to recognize habitual language patterns that may otherwise have gone unnoticed. For example, one clinician described becoming aware of relying on negative warnings and expressed readiness to begin using more neutral language immediately. These findings suggest that the combination of diabetes-specific language guidance, simulation, and structured reflection may be particularly useful for promoting awareness and motivation for communication change.

This study also highlights the added value of a targeted, guideline-linked intervention compared with general communication training. Participants in the active control condition often described the general patient-centered communication video as familiar or more appropriate for learners earlier in training. In contrast, participants who received the stigma-reduction training module described the content as actionable and connected to authoritative recommendations. This distinction suggests that primary care clinicians may be more engaged by communication training when it is specific, evidence-based, and directly relevant to a clinical challenge they encounter in practice.

To our knowledge, this is one of the first educational modules to combine a standardized patient encounter, video self-review, transcript-guided reflection, and theory-informed content to target stigma reduction in clinician-patient communication about T2D. This multimodal design allowed clinicians to observe and reflect on their language in a realistic context while receiving concrete strategies for adopting person-first, strengths-based alternatives. Prior research using simulated encounters and video-based reflection supports the value of self-review as a nonthreatening approach for helping learners identify opportunities for improvement [34]. This was particularly relevant to the present module, in which standardized patients did not provide direct feedback, preserving the authenticity of the encounter and allowing clinicians to identify their own opportunities for language change.

Participants also identified implementation challenges, including time constraints and the need for practical language tools. These findings underscore the importance of developing ready-to-use resources that can be integrated into routine primary care workflows. Because both study conditions were similar in duration, the greater emphasis on time constraints among active control participants may reflect differences in perceived relevance or motivation rather than time burden alone. Nevertheless, time pressure is a real and persistent feature of primary care, and future implementation efforts should address both clinician engagement and workflow feasibility.

XSL•FO
RenderX

Finally, the findings are consistent with habit-breaking approaches to behavior change, which emphasize awareness, motivation, concrete alternative strategies, and continued practice [35]. The stigma-reduction training module appeared to address several of these elements by increasing awareness of language use, motivating clinicians to reconsider familiar terminology, and providing actionable alternatives. However, the single-session format limited opportunities for repeated practice and reinforcement [36]. Future studies should examine whether booster sessions, reminders, longitudinal reflection, or integration into continuing medical education can support sustained changes in clinician language use in real-world practice.

Limitations and Future Directions

This pilot was limited by its small sample size and potential self-selection bias, as clinicians with a greater interest in communication quality may have been more likely to participate. The simulated clinical environment may not fully capture the complexity of real-world primary care practice, and the brief follow-up window precluded evaluation of changes in self-reported clinician behavior. In addition, the study did not assess postintervention changes in perceived behavioral control. Future studies should assess whether changes in clinicians' attitudes and intentions are sustained over time and whether they translate into changes in clinician language use and patient experiences.

The small sample size also limits precision in estimating effect sizes. Although baseline attitude scores did not differ statistically between groups, the stigma-reduction training module group had lower baseline attitude scores than the active control condition, creating greater room for improvement and raising the possibility that regression to the mean contributed to this observed change. Consistent with the stage model of behavioral therapies research, this study should therefore be interpreted as an early-stage pilot trial intended to evaluate acceptability, feasibility, and preliminary outcome patterns rather than definitive efficacy [33]. The large observed effect on attitudes should be interpreted with caution, but it provides an empirical signal supporting a more definitive, powered randomized trial.

Recruitment was protracted, spanning over 12 months to enroll 24 clinicians. This raises feasibility considerations for future trials and is consistent with prior literature documenting challenges in recruiting community-based primary care clinicians [37]. The high participation burden of this pilot, which required an in-person, recorded standardized patient encounter, likely contributed to the extended recruitment timeline [38]. Future studies should consider more flexible recruitment and participation strategies, including remote or hybrid formats.

Although qualitative analysts were not provided with study condition labels, full blinding was not possible because verbatim exit interview transcripts may have included comments that revealed the condition received.

Conclusions

In summary, this mixed methods pilot study suggests that a brief, multicomponent, theory-informed stigma-reduction training module is feasible, acceptable, and associated with improved primary care clinicians' attitudes toward avoiding stigmatizing language in T2D care. Changes in intentions were descriptively in the expected direction but did not reach statistical significance. Given its strong acceptability, this module may hold promise for integration into continuing medical education and professional development. Future studies should assess long-term effects, feasibility in real-world clinical settings, changes in clinician language use, and impact on patient experiences and outcomes, with the goal of advancing more compassionate and equitable diabetes care.

Acknowledgments

The authors gratefully acknowledge Michael Ramus and Agnes Dunne for their assistance as standardized patient actors, as well as the many research assistants and associates involved in planning and execution of the project, including Natalie Leonard, Anderson Riley, Francis Marrero Rosado, and Madison Taylor. The authors also thank Briana Mezuk, PhD, and Mary de Groot, PhD, for their mentorship and guidance, and Lise Anderson for her professional editorial support.

The authors used the University of Michigan's internal ChatGPT (GPT-5.2; OpenAI) service to suggest language improvements in the manuscript. The authors reviewed and edited the AI-assisted output and take full responsibility for the content of the manuscript. The authors also used Claude Sonnet 4.5 (Anthropic) to support statistical analysis by helping to identify an appropriate effect-size package in R, troubleshoot R code, and format the presentation of statistical results. All statistical decisions, code, and outputs were reviewed by members of the analytic team, including AC and RG, who have expertise in statistical analysis.

Funding

This research was funded by an American Diabetes Association pilot grant (CDTR-03).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: KLJ (lead), AA (supporting), AC (supporting), NH (supporting), RG (supporting)
Data curation: AC (lead), BPL (equal), DW (equal)

XSL•FO
RenderX

Formal analysis: KLJ (lead), AC (equal), AA (supporting), BD (supporting), SL (supporting), DW (supporting), HT (supporting), MR (supporting), RG (supporting)
Funding acquisition: KLJ (lead), AA (supporting), AC (supporting), NH (supporting), RG (supporting)
Investigation: AC (lead), DW (equal), BPL (equal), BD (supporting), SL (supporting), HT (supporting), MR (supporting), RG (supporting)
Methodology: KLJ (lead), AC (equal), AA (supporting), BPL (supporting)
Project administration: AC (lead), DW (equal), BPL (equal), BD (supporting), SL (supporting), HT (supporting), MR (supporting), RG (supporting)
Resources: KLJ (lead), AA (equal), RG (equal)
Supervision: KLJ (lead), AC (equal), DW (equal)
Validation: KLJ (lead), AC (equal), DW (equal), RG (equal)
Visualization: KLJ (lead), AC (equal), DW (equal)
Writing – original draft: KLJ (lead), AC (equal), DW (supporting), BD (supporting), SL (supporting), MR (supporting)
Writing – review & editing: KLJ (lead), AC (equal), NH (supporting), RG (supporting)

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was not registered, explained by authors as follows: pilot study. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the study was considered formative, guiding the development of an application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Study visit protocol and choreography.
[PDF File, 180 KB - diabetes_v11i1e95805_app1.pdf]

Multimedia Appendix 2
Guided self-reflection activity.
[PDF File, 167 KB - diabetes_v11i1e95805_app2.pdf]

Multimedia Appendix 3
Theory of planned behavior survey.
[PDF File, 253 KB - diabetes_v11i1e95805_app3.pdf]

Multimedia Appendix 4
Acceptability survey.
[PDF File, 189 KB - diabetes_v11i1e95805_app4.pdf]

Multimedia Appendix 5
Semistructured exit interview.
[PDF File, 166 KB - diabetes_v11i1e95805_app5.pdf]

Multimedia Appendix 6
Sensitivity analyses by analytic approach.
[PDF File, 101 KB - diabetes_v11i1e95805_app6.pdf]

Checklist 1
CONSORT checklist.
[PDF File, 196 KB - diabetes_v11i1e95805_app7.pdf]

References

1. National diabetes statistics report. Centers for Disease Control and Prevention (CDC). URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2026-01-09]

XSL•FO
RenderX

2. Khunti K, Gomes MB, Pocock S, et al. Therapeutic inertia in the treatment of hyperglycaemia in patients with type 2 diabetes: a systematic review. Diabetes Obes Metab 2018 Feb;20(2):427-437. [doi: 10.1111/dom.13088] [Medline: 28834075]
3. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social determinants of health and diabetes: a scientific review. Diabetes Care 2020 Nov 2;44(1):258-279. [doi: 10.2337/dci20-0053] [Medline: 33139407]
4. Browne JL, Ventura A, Mosely K, Speight J. 'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes. BMJ Open 2013 Nov 18;3(11):e003384. [doi: 10.1136/bmjopen-2013-003384] [Medline: 24247325]
5. Speight J, Skinner TC, Dunning T, et al. Our language matters: improving communication with and about people with diabetes. A position statement by Diabetes Australia. Diabetes Res Clin Pract 2021 Mar;173:108655. [doi: 10.1016/j.diabres.2021.108655] [Medline: 33422586]
6. Link BG, Phelan JC. Conceptualizing Stigma. Annu Rev Sociol 2001 Aug;27(1):363-385. [doi: 10.1146/annurev.soc.27.1.363]
7. Speight J, Holmes-Truscott E, Garza M, et al. Bringing an end to diabetes stigma and discrimination: an international consensus statement on evidence and recommendations. Lancet Diabetes Endocrinol 2024 Jan;12(1):61-82. [doi: 10.1016/S2213-8587(23)00347-9] [Medline: 38128969]
8. Holmes-Truscott E, Browne JL, Ventura AD, Pouwer F, Speight J. Diabetes stigma is associated with negative treatment appraisals among adults with insulin-treated type 2 diabetes: results from the second Diabetes MILES - Australia (MILES-2) survey. Diabet Med 2018 May;35(5):658-662. [doi: 10.1111/dme.13598] [Medline: 29417612]
9. Puhl RM, Himmelstein MS, Hateley-Browne JL, Speight J. Weight stigma and diabetes stigma in U.S. adults with type 2 diabetes: associations with diabetes self-care behaviors and perceptions of health care. Diabetes Res Clin Pract 2020 Oct;168:108387. [doi: 10.1016/j.diabres.2020.108387] [Medline: 32858100]
10. Sims TJ, Chinthammit C, Constantine ML, Bushnell DM, Spaepen E. Examining the relationship between weight stigma, diabetes stigma, and HbA1c in adults with type 2 diabetes. Diabetes Ther 2025 Nov;16(11):2137-2156. [doi: 10.1007/s13300-025-01795-8] [Medline: 40974396]
11. Holmes-Truscott E, Ekpor E, Litterbach E, et al. From language matters to stigma-free communication in diabetes: evolution, challenges, and next steps. Diabet Med 2026 Apr 5:e70311. [doi: 10.1111/dme.70311] [Medline: 41937302]
12. NIMHD research framework. National Institute on Minority Health and Health Disparities (NIMHD). 2025. URL: https://www.nimhd.nih.gov/resources/nimhd-research-framework [accessed 2026-01-09]
13. American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S27-S49. [doi: 10.2337/dc25-S002] [Medline: 39651986]
14. Pérez-Stable EJ, El-Toukhy S. Communicating with diverse patients: how patient and clinician factors affect disparities. Patient Educ Couns 2018 Dec;101(12):2186-2194. [doi: 10.1016/j.pec.2018.08.021] [Medline: 30146407]
15. American Diabetes Association Professional Practice Committee for Diabetes. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S89-S131. [doi: 10.2337/dc26-S005] [Medline: 41358898]
16. Bannuru RR, Prieto F, Murdock L, Tollefson E. Diabetes management: a case study to drive national policy change in primary care settings. Ann Fam Med 2024;22(6):550-556. [doi: 10.1370/afm.3175] [Medline: 39586698]
17. Dickinson JK, Guzman SJ, Maryniuk MD, et al. The use of language in diabetes care and education. Diabetes Care 2017 Dec;40(12):1790-1799. [doi: 10.2337/dci17-0041] [Medline: 29042412]
18. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
19. Nyblade L, Stockton MA, Giger K, et al. Stigma in health facilities: why it matters and how we can change it. BMC Med 2019 Feb 15;17(1):25. [doi: 10.1186/s12916-019-1256-2] [Medline: 30764806]
20. Stangl AL, Earnshaw VA, Logie CH, et al. The Health Stigma and Discrimination Framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. BMC Med 2019 Feb 15;17(1):31. [doi: 10.1186/s12916-019-1271-3] [Medline: 30764826]
21. Talumaa B, Brown A, Batterham RL, Kalea AZ. Effective strategies in ending weight stigma in healthcare. Obes Rev 2022 Oct;23(10):e13494. [doi: 10.1111/obr.13494] [Medline: 35934011]
22. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
23. Health Research University of Michigan. URL: https://umhealthresearch.org/ [accessed 2026-01-09]
24. University of Michigan HomeLab Core Facility (RRID:SCR_026753). RRID Portal. URL: https://rrid.site/data/record/nlx_144509-1/SCR_026753/resolver [accessed 2026-06-19]
25. Joiner K, Agapiou A, Dickinson JK, Adams M, Piatt G. An educational video to promote the use of stigma-free language by primary care clinicians in interactions with adults with type 2 diabetes: a qualitative study. Prim Health Care Res Dev 2025 Dec 29;27:e5. [doi: 10.1017/S1463423625100728] [Medline: 41457429]
26. Communication skills: a patient-centered approach. wmumed YouTube page. 2015 Jul 9. URL: https://www.youtube.com/watch?v=S4wWClQhZaA [accessed 2026-01-09]
27. Paskins Z, McHugh G, Hassell AB. Getting under the skin of the primary care consultation using video stimulated recall: a systematic review. BMC Med Res Methodol 2014 Aug 30;14:101. [doi: 10.1186/1471-2288-14-101] [Medline: 25175450]

XSL•FO
RenderX

28. Paskins Z, Sanders T, Croft PR, Hassell AB. Exploring the added value of video-stimulated recall in researching the primary care doctor–patient consultation. Int J Qual Methods 2017 Dec 1;16(1). [doi: 10.1177/1609406917719623]
29. Lyddon WJ, Yowell DR, Hermans HJM. The self-confrontation method: theory, research, and practical utility. Couns Psychol Q 2006 Mar;19(1):27-43. [doi: 10.1080/09515070600589719]
30. Ajzen I. Constructing a theory of planned behavior questionnaire. : University of Massachusetts Amherst; 2006 URL: https://people.umass.edu/aizen/pdf/tpb.measurement.pdf [accessed 2026-01-09]
31. Weymiller AJ, Montori VM, Jones LA, et al. Helping patients with type 2 diabetes mellitus make treatment decisions: statin choice randomized trial. Arch Intern Med 2007 May 28;167(10):1076-1082. [doi: 10.1001/archinte.167.10.1076] [Medline: 17533211]
32. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
33. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Sci Pract 2001;8(2):133-142. [doi: 10.1093/clipsy.8.2.133]
34. Giles AK, Carson NE, Breland HL, Coker-Bolt P, Bowman PJ. Use of simulated patients and reflective video analysis to assess occupational therapy students' preparedness for fieldwork. Am J Occup Ther 2014;68 Suppl 2(Suppl 2):S57-S66. [doi: 10.5014/ajot.2014.685S03] [Medline: 25397940]
35. Hagiwara N, Kron FW, Scerbo MW, Watson GS. A call for grounding implicit bias training in clinical and translational frameworks. Lancet 2020 May 2;395(10234):1457-1460. [doi: 10.1016/S0140-6736(20)30846-1] [Medline: 32359460]
36. Devine PG, Forscher PS, Austin AJ, Cox WTL. Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. J Exp Soc Psychol 2012 Nov;48(6):1267-1278. [doi: 10.1016/j.jesp.2012.06.003] [Medline: 23524616]
37. Johnston S, Liddy C, Hogg W, Donskov M, Russell G, Gyorfi-Dyke E. Barriers and facilitators to recruitment of physicians and practices for primary care health services research at one centre. BMC Med Res Methodol 2010 Dec 13;10:109. [doi: 10.1186/1471-2288-10-109] [Medline: 21144048]
38. Asch S, Connor SE, Hamilton EG, Fox SA. Problems in recruiting community-based physicians for health services research. J Gen Intern Med 2000 Aug;15(8):591-599. [doi: 10.1046/j.1525-1497.2000.02329.x] [Medline: 10940152]

Abbreviations

ADA: American Diabetes Association
ADCES: Association of Diabetes Care & Education Specialists
CALI: computer-aided lab interviewing
CONSORT: Consolidated Standards of Reporting Trials
T2D: type 2 diabetes

Edited by S Li; submitted 21.Mar.2026; peer-reviewed by N Keesmekers, Q Fan; revised version received 18.Jul.2026; accepted 22.Jul.2026; published 11.Aug.2026.

Please cite as:
Joiner KL, Agapiou A, Carmichael A, Labbree BP, Walter D, Tarraf H, Raju M, Dogari B, Hagiwara N, Lin S, Gonzalez R
Promoting Stigma-Free Language in Type 2 Diabetes Care Among Primary Care Clinicians: Randomized Controlled Pilot Trial
JMIR Diabetes 2026;11:e95805
URL: https://diabetes.jmir.org/2026/1/e95805
doi:10.2196/95805

© Kevin L Joiner, Alexandra Agapiou, Alicia Carmichael, Brandon P Labbree, Donna Walter, Houda Tarraf, Medha Raju, Boluwatife Dogari, Nao Hagiwara, Sophia Lin, Richard Gonzalez. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Content Validation of an Electronic Health Record–Based Diabetes Self-Management Support Tool for Older Adults With Type 2 Diabetes: Qualitative Study

Ploypun Narindrarangkura[1], MS, MD, PhD; Siroj Dejhansathit[2,3], MD; Uzma Khan[2,3], MD; Margaret Day[4], MSPH, MD; Suzanne A Boren[5,6], MHA, PhD; Eduardo J Simoes[5,7], MS, MPH, MD; Min Soon Kim[5,6], PhD

[1]Phramongkutklao College of Medicine, Bangkok, Thailand
[2]Cosmopolitan International Diabetes and Endocrinology Center, University of Missouri, Columbia, MO, United States
[3]Department of Medicine, University of Missouri, Columbia, MO, United States
[4]Department of Family and Community Medicine, University of Missouri, Columbia, MO, United States
[5]University of Missouri Institute for Data Science and Informatics, Columbia, MO, United States
[6]Department of Health Sciences, College of Health Sciences, University of Missouri, 310 Clark Hall, Columbia, MO, United States
[7]Department of Biomedical Informatics, Biostatistics and Medical Epidemiology, University of Missouri, Columbia, MO, United States

Corresponding Author:
Min Soon Kim, PhD
University of Missouri Institute for Data Science and Informatics, Columbia, MO, United States

Abstract

Background: Older adults with diabetes frequently access their electronic health record (EHR) notes but often report difficulty understanding medical jargon and nonspecific self-care instructions. To address this communication gap, we developed Support-Engage-Empower-Diabetes (SEE-Diabetes), a patient-centered, EHR-integrated diabetes self-management support tool designed to embed tailored educational statements within the assessment and plan section of clinical notes.

Objective: This study aimed to validate the clarity, relevance, and alignment of SEE-Diabetes content with the Association of Diabetes Care & Education Specialists 7 Self-Care Behaviors framework from the perspectives of older adults and clinicians.

Methods: An interdisciplinary team conducted expert reviews and qualitative interviews with 11 older adults with diabetes and 8 clinicians practicing in primary care (family medicine) and specialty diabetes care settings at a Midwestern academic health center. Patients evaluated the readability and relevance of the content, while clinicians assessed clarity, sufficiency, and potential clinical utility. Interview data were analyzed using inductive thematic analysis, and descriptive statistics were used to summarize participant characteristics.

Results: Patients (mean age 72, SD 4.9 y; mean diabetes duration 26, SD 15 y) reported that the SEE-Diabetes statements were clear, relevant, and written in plain language that supported understanding of self-care recommendations. Clinicians (mean 13, SD 9.5 y of diabetes care experience) viewed the content as concise, clinically appropriate, and well aligned with patient self-management goals and the Association of Diabetes Care & Education Specialists 7 Self-Care Behaviors framework. Both groups identified the tool's potential to enhance patient engagement and patient-clinician communication, while noting opportunities to improve the specificity of language, particularly within medication-related content.

Conclusions: SEE-Diabetes demonstrated content validity as a practical, patient-centered digital health tool for supporting diabetes self-management communication within EHR clinical notes. The findings support its use as a complementary approach to reinforce self-care communication in routine clinical practice and highlight areas for refinement to enhance personalization.

(JMIR Diabetes 2026;11:e83448) doi:10.2196/83448

KEYWORDS

diabetes mellitus, type 2; electronic health records; self-management; patient education; older adults; digital health; health literacy

Introduction

Background

Diabetes is highly prevalent among older adults in the United States, with an estimated 29.2% of adults aged 65 years or older having been diagnosed with or undiagnosed diabetes during 2017‑2020, and approximately 48.8% of adults in this age group had prediabetes according to the most recent National Diabetes Statistics Report [1]. As the aging population grows, primary care clinicians face increasing pressure to deliver effective, individualized diabetes self-management education

XSL•FO
RenderX

within routine visits. Diabetes self-management education and support (DSMES) has been shown to improve glycemic control, reduce complications, and enhance self-efficacy [2-5]. However, the delivery of DSMES in outpatient settings is frequently constrained by limited visit time, complex documentation requirements, challenges in referral and access, and poor integration with routine clinical workflows [6,7].

National DSMES standards outline 4 critical times when individuals with diabetes should receive structured education and support [2]; however, referrals and access to formal DSMES services remain inconsistent. As a result, self-management guidance is often delivered informally during routine visits, underscoring the need for tools that reinforce evidence-based messaging within existing clinical workflows.

To address these challenges, our team developed Support-Engage-Empower-Diabetes (SEE-Diabetes), a patient-centered educational aid designed to support clinicians in delivering tailored diabetes education to older adults during clinic visits. SEE-Diabetes integrates directly into the electronic health record (EHR) by embedding brief, personalized education statements—drawn from a curated content library—into the assessment and plan section of the clinician's note. The content is organized according to the 7 core domains of the Association of Diabetes Care & Education Specialists 7 Self-Care Behaviors (ADCES7), including healthy coping, healthy eating, being active, taking medication, monitoring, reducing risk, and problem solving [8].

Placement of SEE-Diabetes in the Assessment and Plan section was intentional. Prior formative research with older adults with diabetes from our group found that the majority (80%) accessed and read their clinic notes through patient portals, yet many found these notes difficult to understand due to medical jargon and vague or nonactionable self-care guidance [6,7]. Embedding clear, relevant, and actionable statements in a section that patients already read may therefore address an important communication gap while also integrating seamlessly into clinician documentation.

SEE-Diabetes was developed using a user-centered design (UCD) approach to ensure alignment with real-world clinical needs [9,10]. The first stage of development involved an analysis of EHR documentation patterns related to diabetes care [11], followed by a second stage comprising focus groups with older adults with type 2 diabetes and clinicians involved in diabetes management to identify gaps in the clarity, readability, and consistency of self-management information [6,7]. This study represents the third stage of the UCD process and focuses on content validation of the SEE-Diabetes educational statements to ensure their accuracy, relevance, and practical utility for both patients and clinicians [12].

Objective

Our objective was to assess the clarity, helpfulness, and perceived value of SEE-Diabetes education content by conducting in-depth interviews with older adults and clinicians practicing in primary care (family medicine) and specialty diabetes care settings. This validation step is essential before the broader implementation of SEE-Diabetes in primary care settings. By embedding actionable, comprehensible diabetes education into clinical notes, SEE-Diabetes may enhance patient understanding, improve continuity of care, and support more effective chronic disease management among older adults.

Methods

Study Design

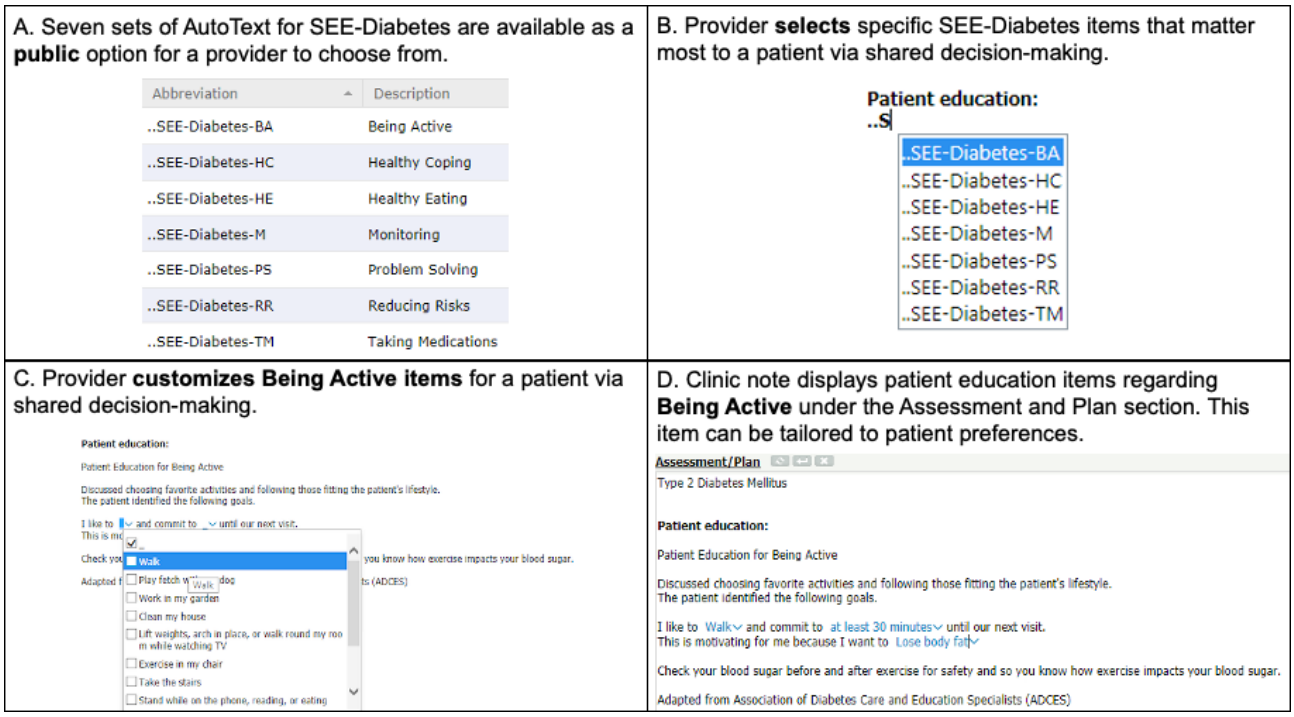
We conducted a qualitative content validation study to assess the clarity, readability, and clinical relevance of SEE-Diabetes, an EHR-integrated education tool for older adults with diabetes. This phase represented the third stage of a UCD process. The content validation process included (1) expert reviews by clinicians and certified diabetes care and education specialists, and (2) user feedback through semistructured interviews with older adults with diabetes and with primary care or endocrinology clinicians. The interdisciplinary research team included experts in informatics, endocrinology, primary care, and diabetes education.

Description of SEE-Diabetes

SEE-Diabetes content was implemented within the EHR as "auto-text" templates in Oracle Cerner's PowerChart. During documentation, the clinician first selects the SEE-Diabetes category most relevant to the patient's needs, informed by shared decision-making during the visit. Within the category chosen, the clinician can review and select multiple educational statements addressing specific self-care behaviors. Each statement can be further customized to reflect the patient's individual preferences, goals, literacy level, and clinical circumstances. Examples of customization include changing the activity type (eg, walking and gardening) or specifying behavior targets (eg, number of minutes per day) (Figure 1).

XSL•FO
RenderX

Figure 1. Overview of the Support-Engage-Empower-Diabetes framework illustrating integration of tailored patient education statements into electronic health records, aligned with the Association of Diabetes Care & Education Specialists 7 Self-Care Behaviors. (A) Seven publicly available autotext sets (being active, healthy coping, healthy eating, monitoring, problem solving, reducing risks, and taking medications) are mapped to Association of Diabetes Care & Education Specialists 7 Self-Care Behavior domains. (B) The clinician selects the relevant Support-Engage-Empower-Diabetes category in the Patient education field. (C) Within the chosen category, statements are customized collaboratively (eg, activity type, frequency, or targets) during shared decision-making. (D) The finalized, tailored patient education statements are inserted into the Assessment and Plan section of the clinic note and become available to patients via the portal. ADCES7: Association of Diabetes Care & Education Specialists 7 Self-Care Behavior; SEE-Diabetes: Support-Engage-Empower-Diabetes.



The finalized patient education text was embedded within the assessment and plan section of the clinic note. Embedding SEE-Diabetes in the assessment and plan positions the guidance where patients already expect to find follow-up instructions, while requiring minimal change to clinician workflow. Insertion and customization generally take less than 1 minute, minimizing any disruption to the visit flow.

Study Setting

The study was conducted at the University of Missouri Health Care, an academic medical center serving 114 counties in Missouri [13]. The center uses the Oracle Cerner PowerChart EHR system to consolidate patient data across facilities. Patients can access their medical records, including clinic notes, through the HEALTHConnect portal. Clinic notes were retrieved from PowerChart, and patient recruitment was facilitated using PowerInsight, Oracle Cerner's operational reporting platform.

Interview Development

A total of 3 representative clinical scenarios were developed based on de-identified data from older adults with type 2 diabetes. For each case, SEE-Diabetes was applied to generate tailored patient education statements aligned with ADCES7 domains. The scenarios were (1) a 69-year-old woman with uncontrolled diabetes (monitoring and healthy eating), (2) a 72-year-old man with stable diabetes and obesity (medication adherence and risk reduction), and (3) a 67-year-old woman with type 2 diabetes (physical activity and healthy coping). An endocrinologist drafted the clinic notes (history of present illness and assessment and plan), and the multidisciplinary team reviewed all content for clinical accuracy and guideline concordance. The history of present illness sections are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. History of present illness section of three clinic notes. These are in-model screenshots of three example clinic notes designed for patients with diabetes aged 65 years and older attending follow-up visits at the Cosmopolitan International Diabetes and Endocrinology Center. Participants were asked to review the history of present illness section along with the assessment and plan.

Clinic Note 1: History of Present Illness

- 69-year-old lady presents for discussion regarding long-term management of diabetes mellitus.
- Initially diagnosed in 2009,
- started on metformin 1000 mg BID
- glimepiride 4 mg BID started in 2010
- pioglitazone 45 mg QD in the morning started in 2019
- has never been on insulin
- She has not had any diabetic education since her diagnosis. Denies numbness/tingling in her extremities. She has tried a keto diet in the past, but this led to frequent hypoglycemia. Since July 2022, she has been eating <1700 calories daily, which resulted in weight gain. Eye exam was done in April 2022 but not sure if her eyes were dilated. She has noted that her vision changes as her BG fluctuates, and sometimes her vision is blurry despite wearing her bifocals. There is no family history of T2DM, T1DM, or osteoporosis, and has never had a DEXA scan. She sees gynecologist yearly but does not have a regular well-woman exam.

Clinic Note 2: History of Present Illness

- This is a 72-year-old gentleman who presents for follow-up of his diabetes.
- He was diagnosed with diabetes around the age of 50 and has been on metformin since that time.
- Blood glucose is slightly worse; he checks every day and has been running above 150 mg/dl
- He is on Metformin 1000 mg twice a day. He has noted increased blood glucose since he got Covid in 1/2022.
- He has had fatigue, feels nauseous, so has not been taking his metformin daily, only taking it "on good days."
- He was supposed to meet with a dietitian, but has had so many doctors' appointments that did not make it.
- His family doctor added a "small pill every day," but he is not sure what medication it is.
- He does not want to use any medications that are injections at this time and feels he can control his diabetes once he is feeling better.
- His $HbA1c_{1c}$ was 7.5%; it has increased now to 8.8%.
- He plans on focusing on lifestyle and has been having increased burning in feet, so he was not walking much.

Clinic Note 3: History of present illness

- 67-year-old patient who presents to discuss diabetes mellitus type 2 management
- DMT2 diagnosed age 55 years. There is no retinopathy, neuropathy; she has microalbuminuria. She also has hyperlipidemia and hypertension
- Current regimen includes glipizide 10 mg daily and metformin extended release 500 mg, takes 1 tablet twice a day
- Her last diabetes class was before 2014
- Checks FSG once a day, ranges from 122-140s, no hypoglycemia
- She was walking, had to quit because of arthritis, now spends most of her time at home, and feels discouraged about her diabetes
- She likes to bake but has no motivation to do it anymore. Three friends have passed away in the last four years, and she has no family near home. She tries to eat healthy, mainly frozen meals.
- She is a smoker, has been trying to quit but feels she cannot do it.
- Blood pressure had been controlled on triamterene/HCTZ, 37.5/25 mg, and losartan 100 mg daily, but has increased and now also amlodipine 10 mg daily. For hyperlipidemia, takes pravastatin 10 mg daily.
- Takes ASA 81 mg daily.
- Last eye exam was on May 5, 2022, and showed no retinopathy. She has had cataract surgery also.
- Last urine microalbumin on October 14, 2022, showed microalbuminuria (high: 93)
- Denies numbness in feet or tingling, no foot ulcers
- No chest pain, palpitations, no nausea, vomiting, diarrhea, no abdominal pain, no cough or fever

Parallel semistructured interview guides were developed for patients and clinicians. Patient interviews consisted of 4 open-ended questions assessing readability, helpfulness, relevance, and anticipated future use of the SEE-Diabetes

XSL•FO
RenderX

statements. Clinicians reviewed the same notes and answered 4 corresponding questions addressing clarity, completeness, clinical applicability, and suggestions for improvement. This mirrored design enabled direct comparison of perspectives across patient and clinician groups.

Data Collection and Analysis

Participants were recruited from Family and Community Medicine clinics and the Cosmopolitan International Diabetes and Endocrinology Center in October-November 2022. Participants were asked to evaluate the Patient Education section generated via SEE-Diabetes, which was included under the assessment and plan section of the 3 clinic notes for patients (Textbox 2).

XSL•FO
RenderX

Textbox 2. Assessment and plan section of three clinic notes. These are in-model screenshots of the Assessment and plan sections from three example clinic notes for patients with diabetes aged 65 years and older. The patient education sections were generated using Support-Engage-Empower-Diabetes, based on reviews of each patient, and then customized by an endocrinologist. Subsequently, they were reviewed by other team members. Participants were asked to review the Patient Education section and answer open-ended questions to assess the readability, helpfulness, and values of Support-Engage-Empower-Diabetes.

Clinic Note 1 : Assessment and Plan

1. Uncontrolled type 2 diabetes mellitus with hyperglycemia

- Reviewed lab results with patient emphasizing the importance of optimizing HbA_{1c}, with target below 8%
- Advised to check FSG regularly and record, bring records for review next visit
- Reviewed risks of hypoglycemia, prevention, and management of hypoglycemic episodes
- Reviewed foot care, call me if notice an open area on foot
- She will schedule an eye exam

Patient Education for Monitoring

- Monitoring is an important aspect of self-care. It helps you know if you are meeting recommended treatment goals to keep you healthy.
- My goal is to learn how to use my monitor, learn how to interpret my blood sugar levels
- I want to use this information to learn how different foods affect my blood sugar
- I commit to checking my blood sugar at the following times: 1 time a day and plan to bring in my readings to my next visit

Adapted from Association of Diabetes Care and Education Specialists (ADCES)

2. Obesity

- The patient is motivated to use weight control, which will improve metabolic health, including diabetes mellitus type 2, hypertension, and hyperlipidemia.

Patient Education for Healthy Eating

Discussed the meal plan today and the patient set the following goals:

- I will read the Nutrition Facts Label.
- I will add 2 servings of vegetables to my diet.
- I will cut down added sugar in my drinks from my diet to help to control my blood sugar.
- I plan to learn more about considering different healthy eating options by meeting with a diabetes specialist by the time of our next visit.

Adapted from Association of Diabetes Care and Education Specialists (ADCES)

Clinic Note 2: Assessment and Plan

1. Type 2 diabetes mellitus without complications

- Reviewed lab results with patient emphasizing the importance of optimizing HbA_{1c}, with target HbA_{1c} below 8 %
- Advised to check FSG regularly and record, bring records for review next visit,
- Reviewed risks of hypoglycemia, prevention, and management of hypoglycemic episodes
- Reviewed foot care, call me if notice an open area on foot

Patient Education for Taking Medications

- Taking medications helps lower your risk for heart attack, stroke, and kidney damage by managing blood glucose, blood pressure, and cholesterol levels in your body. The longer you have diabetes, the more help you will need from medications to keep you and your heart, eyes, and kidneys healthy.
- I plan to take my medications on time by bringing in all my medications to my next appointment between now and my next visit.

Adapted from Association of Diabetes Care and Education Specialists (ADCES)

2. Body mass index 40+ - severely obese (finding)

- Patient has started to feel somewhat better after his COVID infection and is motivated to increase activity and control his weight to improve management of his diabetes, hyperlipidemia.

Patient Education for Reducing Risks

XSL•FO
RenderX

- Reducing risks means doing behaviors that minimize or prevent complications and negative outcomes of prediabetes and diabetes. Risks mean doing behaviors that minimize or prevent complications and negative outcomes of prediabetes and diabetes.
- I plan to make positive lifestyle changes, participate in diabetes self-management education.
- I will do this by scheduling an appointment by the time of our next visit.

Adapted from the Association of Diabetes Care and Education Specialists (ADCES)

- Follow-up in clinic in 3 months with labs before the appointment
- Referral placed for diabetes education again

Clinic Note 3: Assessment and Plan

1. Diabetes Mellitus

- Detailed discussion with the patient, reviewed HbA_{1c} of 7.2%, her target is below 8% so she is doing well. HbA_{1c} of 7.2%, her target is below 8% so she is doing well.
- However, she has gained weight and is not feeling well.
- We discussed medications that might make her mood better; however, the patient wants to focus on positive thinking first.

Patient Education for Being Active

- Discussed choosing favorite activities and following those fitting the patient's lifestyle. The patient identified the following goals.
- I like to walk, park farther away from the door and commit to 10 minutes daily until our next visit. This is motivating for me because I want to improve mood
- Check your blood sugar before and after exercise for safety and so you know how exercise impacts your blood sugar.

Patient Education for Healthy Coping

- Discussed with patient that it is important to find healthy ways to cope and not to turn to harmful habits such as smoking, overeating, drinking or alcohol. This is especially true if you have diabetes. Having a lot of stress can increase blood glucose (sugar) levels, make you feel more negative and may lead to less healthy choices.
- I plan to cope with stress by make a list of people I can turn to for support and report back at my next visit to share how that went. I will observe/record my mood daily, I will seek help if I feel challenged.

Adapted from Association of Diabetes Care and Education Specialists (ADCES)

- She will continue her current medications, focus on lifestyle and I will see her back in 3 months with labs before the appointment. She will call if she needs to make an earlier appointment.

In-depth interviews were conducted in private settings and lasted approximately 30 minutes. Sessions were audio recorded, transcribed verbatim, and de-identified. Descriptive statistics summarized participant demographics. Thematic analysis [14] was conducted using an inductive approach to identify key themes, and transcripts were coded independently by 2 researchers (PN and SD) before being reviewed by the research team.

Ethical Considerations

This study was reviewed and approved by the University of Missouri Health Care Institutional Review Board (IRB #2078424 MU). The protocol was deemed to be no greater than minimal risk. Written informed consent was obtained from all participants, including disclosure of the study goals. Participants could opt out at any time. Nonessential identifying information has been removed for publication. Screenshots and examples included in the manuscript were deidentified so that no individual could be identified directly or indirectly. Participants were compensated with a US $50 cash card.

Results

Patient Characteristics and Thematic Analysis Findings From Interviews

Patient Characteristics

Overall, 11 patients participated, recruited from a specialty diabetes center. The average age was 72 (SD 4.9; range 66 - 83) years, 6 were female (55%), and most were non-Hispanic White (10/11, 91%). Nearly half (5/11, 45.5%) had some college education. The mean duration of diabetes was 26 (SD 15; range 3 - 47) years, with a mean hemoglobin A1c (HbA_{1C}) of 7.6% (SD 1.2%; range 6.1% - 10.3%). Most patients were insulin users (9/11, 82%) and routinely accessed their clinic notes via patient portals (10/11, 91%), typically on their own computers (Table 1).

XSL•FO
RenderX

Table . Characteristics of patient participants (n=11).

Characteristics	Values, n (%)
Clinic location	
Cosmopolitan International Diabetes and Endocrinology Center	11 (100)
Age (years), mean (SD; range)	71.6 (4.9; 66-83)
Sex	
Male	5 (45.5)
Female	6 (54.5)
Hispanic or Latino	
No	11 (100)
Yes	0 (0)
Race	
Non-Hispanic White	10 (90.9)
Asian	1 (9.1)
Education	
Some college credit, no degree	5 (45.5)
Associate degree	2 (18.2)
High school graduate, diploma, or equivalent	1 (9.1)
Bachelor's degree	1 (9.1)
Trade/technical/vocational training	1 (9.1)
Higher than a bachelor's degree	1 (9.1)
Diabetes duration (years), mean (SD; range)	25.6 (15; 3-47)
HbA_{1c}, mean (SD; range)	7.6 (1.2; 6.1 - 10.3)
Insulin	
No	2 (18.2)
Yes	9 (81.8)
Access patient portal	
No	1 (9.1)
Yes	10 (90.9)
How (n=10)	
Yourself	9 (90)
With help from someone else	1 (10)
Devices (n=10)	
Computer	8 (80)
I appreciate the large screen (n=2)	—[a]
It's easy (n=2)	—
Mobile devices	2 (20)
My phone is always with me (n=1)	—
Read clinic notes	
No	1 (9.1)
Yes	10 (90.9)

[a]Not applicable.

XSL•FO
RenderX

Readability

Most participants described the SEE-Diabetes statements as straightforward and easy to read due to plain language and clear structure. For example, a 74-year-old woman (HbA_{1c} 6.9%) highlighted that the section:

> *gives you the information about any testing that you have had and the results from it.*

While an 83-year-old man remarked it was:

> *well written and easily understood.*

However, some participants suggested adopting stronger motivational phrasing that better reflected a patient's voice to encourage action, such as statements:

> [to get them to take something seriously 71-year-old man, HbA1c 6.7%]

Helpfulness

Perceptions of helpfulness were mixed. Several participants valued the content as a practical reminder between visits*:*

> [It makes it a whole lot easier… to remember what I'm supposed to be doing 66-year-old woman, HbA1c 10.3%]

or as a motivator to improve self-care (74-year-old woman, HbA_{1c} 6.9%).

Others, especially those with long-standing diabetes, perceived limited incremental benefit, describing the information as:

> [not new 69-year-old man, HbA1c 9.2%]
>
> [or too broad… not specific enough to make any difference 69-year-old man, HbA1c 9.2%.]

One participant raised concerns about documentation practices, noting frustration with

> [cut and paste… especially when the information is inaccurate 74-year-old woman, HbA1c 7%]

Overall, participants viewed helpfulness as dependent on personalization, specificity, and avoidance of redundant content.

Perceived Value

Patient views on added value also varied. Some appreciated the consolidation of practical information:

> [They don't have to go online and google it. The facts are here 68-year-old woman, HbA1c 6.1%]

and emphasized that SEE-Diabetes could complement physician communication, which was sometimes perceived as incomplete:

> [Doctors aren't the best at communicating all the information. I think those notes actually cover the information… better 69-year-old man, HbA1c 9.2%]

Others reported minimal added value because they were already managing well (76-year-old woman, HbA_{1c} 7.8%) or desired clearer, directive next steps:

> [If there's a diabetes education section… another section with recommendations… I would read that too 71-year-old man, HbA1c 6.7%]

In this context, participants referred to distinct thematic groupings within the SEE-Diabetes content, with actionable recommendations embedded under each of the 7 ADCES7-aligned headings rather than presented in a separate section. Several noted that regular updates and tailoring would be essential to maintain engagement and prevent redundancy. Additional illustrative quotes are provided in Multimedia Appendix 1.

Clinician Characteristics and Thematic Analysis Findings From Interviews

Clinician Characteristics

In total, 8 clinicians participated, including 5 from specialty diabetes care clinics and family medicine (primary care) settings. The average age was 49 (SD 13.5; range 32 - 65) years, and 7 were female (88%). Most were non-Hispanic White (6/8, 75%). The average experience in diabetes care was 13 (SD 12.7; range 2 - 30) years. Most clinicians were familiar with ADCES7 (5/8, 63%) and DSMES guidelines (6/8, 75%) (Table 2).

XSL•FO
RenderX

Table . Characteristics of clinician participants and knowledge of diabetes self-management education and support and Association of Diabetes Care & Education Specialists 7 (n=8).

Characteristics	n (%)
Clinic location	
Cosmopolitan International Diabetes and Endocrinology	5 (62.5)
Keene Family Medicine	2 (25)
Ashland Family Medicine	1 (12.5)
Age (years), mean (SD; range)	48.6 (13.5; 32-65)
Sex	
Male	1 (12.5)
Female	7 (87.5)
Hispanic or Latino	
No	8 (100)
Yes	0 (0)
Race	
Non-Hispanic White	6 (75)
Asian	2 (25)
Work experience (years), mean (SD; range)	12.7 (9.5; 2-30)
Knowledge about DSMES[a] and ADCES7[b] guidelines	
Familiar with ADCES7	
No	3 (37.5)
Yes	5 (62.5)
Familiar DSMES	
No	2 (25)
Yes	6 (75)

[a]DSMES: diabetes self-management education and support.

[b]ADCES7: Association of Diabetes Care & Education Specialists 7 Self-Care Behavior.

Clarity and Concise

Most clinicians agreed that the SEE-Diabetes statements were concise, free of jargon, and written in accessible language. A 40-year-old diabetes specialist noted that the notes “use simple language, no medical jargon, and [are] easy to read.” Similarly, a primary care physician with 2 years’ experience described the information as “short and easy to understand.” However, some clinicians highlighted areas of ambiguity. For instance, a diabetes specialist (8 y experience) observed that the phrasing around medication timing and weight control was confusing and insufficiently specific, suggesting that clearer targets, such as “work on weight loss of 5%,” would enhance patient comprehension.

Sufficiency of Content

Several clinicians endorsed the adequacy of the content, describing it as “pretty thorough and self-explanatory (diabetes specialist, 8 y experience). However, others raised concerns that some sections, particularly related to medication adherence, lacked clarity and risked confusing patients. A primary care physician (2 y experience) noted difficulty interpreting the statement regarding bringing medications to the next appointment, whereas another clinician emphasized the importance of ensuring that each educational category adequately addressed patient priorities.

Clinical Usefulness

Clinicians generally recognized the clinical utility of SEE-Diabetes in supporting patient education and reinforcing self-care. Several reported that the tool aligned with common teaching practices, such as educating patients about blood glucose monitoring, interpreting results, and linking lifestyle behaviors with outcomes (diabetes specialist, 8 y experience). Others saw potential value in emphasizing diabetes-specific goals during visits that are often crowded with competing priorities (primary care physician, 30 y experience). Nonetheless, some cautioned that time constraints may limit consistent use in busy practices. Additionally, suggestions for refinement included offering more concrete examples, such as defining portion sizes in relatable terms (diabetes specialist, 22 y experience), to maximize patient engagement and comprehension. Additional illustrative quotes are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Discussion

Principal Findings

This study validated the content of SEE-Diabetes, an EHR-integrated patient education tool designed to support self-management among older adults with diabetes. By incorporating both expert review and direct feedback from patients and clinicians, we assessed the clarity, relevance, and clinical utility of the educational content. Our findings indicate that SEE-Diabetes has strong potential to address documentation and communication gaps in delivering DSMES and to facilitate more personalized, actionable communication during routine outpatient care. Importantly, SEE-Diabetes is not intended to replace formal DSMES, which remains an ongoing, person-centered process grounded in the assessment of individual learning needs and preferences. Participants may have received varying levels of diabetes education through prior DSMES or routine clinician-provided counseling; however, the amount and modality of such education were not assessed. Accordingly, SEE-Diabetes was evaluated as a complementary, EHR-integrated tool to reinforce routine self-management communication rather than as a measure of DSMES exposure or delivery.

Content validation was conducted using a multimethod approach that combined expert opinion, end-user perspectives, and alignment with the ADCES7 framework [15]. This strategy ensured SEE-Diabetes is grounded in scientific evidence and the practical realities of diabetes care. While content validation is sometimes overlooked in digital health tool development, it plays a critical role in ensuring safety, relevance, and usability. For instance, Patel et al [16] created a clinical decision support system for patients with serious mental illness and diabetes but relied mainly on in silico validation due to the complexity of real-world testing. Such computational methods are useful for assessing technical performance; however, they can delay clinical implementation and may overlook usability issues in practice [17]. In contrast, our study prioritized real-world applicability by engaging both patients and clinicians in the evaluation process, thereby strengthening the credibility and adaptability of SEE-Diabetes in routine care.

Readability and understandability of the educational content emerged as a central theme in the feedback from both patients and clinicians. This aligns with prior evidence that older adults, who may experience cognitive decline or limited health literacy, benefit significantly from materials presented in straightforward, jargon-free language [18]. Communicating health information in clear, familiar terms (eg, using plain language and avoiding medical jargon) significantly improves comprehension and engagement [18]. Participant feedback in our study consistently reinforced the value of plain language in promoting understanding, highlighting the ongoing need for patient-centered communication strategies across health care settings [19]. Ensuring educational content is easily digestible is especially critical for older adults, as it can empower them to more actively participate in their care.

Clinicians viewed SEE-Diabetes as a concise, efficient tool for delivering self-care guidance in time-constrained clinic visits, consistent with prior research showing that brief, targeted educational interventions can be effective in busy health care environments [20-22]. At the same time, some clinicians suggested further refining certain statements (particularly in the "Taking Medication" domain) to enhance clarity and better motivate patients. For example, one provider commented, "Bringing meds to the visit does not ensure the patient will take them regularly between visits." Such feedback underscores the importance of iterative development and continuous user input to ensure that tools like SEE-Diabetes remain clinically relevant, context-sensitive, and adaptable [23]. Incorporating provider and patient suggestions in subsequent revisions will help address these nuances and improve the tool's effectiveness.

Our analysis also identified a remaining gap in the delivery of patient-centered education during routine diabetes follow-up visits. This finding echoes prior studies indicating that although DSMES is widely implemented, it often lacks the personalization necessary to meet individual patient needs [6,7,11,24]. In our previous work, we observed that standard follow-up clinic notes frequently lacked patient-centered education for patients with diabetes [7]. SEE-Diabetes directly addresses this gap by embedding personalized educational content directly into the clinic note (which nearly 80% of our older patients reported reading via the patient portal [7]). By aligning educational messages with each patient's unique context and self-management goals, this approach supports the broader movement toward patient-centered care. Such individualized interventions are expected to enhance patient engagement and treatment adherence and ultimately improve outcomes in diabetes management.

Strengths and Limitations

A key strength of this study lies in its user-centered validation approach, which engaged both patients and clinicians across primary care and specialty care settings. By involving real-world end users in the design and evaluation process, we ensured that SEE-Diabetes content is not only evidence-based but also practical, readable, and clinically relevant. The use of tailored clinical scenarios, combined with in-depth qualitative interviews, provided rich insights into the clarity, usefulness, and perceived value. This multistakeholder engagement enhances the credibility of our findings and supports the tool's adaptability across diverse workflows, thereby strengthening its potential for real-world implementation. Notably, our approach aligns with UCD principles that emphasize iterative development and continuous involvement of target users [9]. By continuously incorporating feedback from both providers and patients, we aimed to develop an educational tool that meets users' needs in everyday practice.

Limitations of this study include a small sample size and a lack of racial and geographic diversity in our participants. Because the majority of participants were non-Hispanic White and recruitment was limited to a single academic health center, the generalizability of our findings may be constrained. This homogeneity is consistent with the demographic profile of the Midwestern United States, where approximately 73% of the population identifies as non-Hispanic White, which likely influenced the composition of our sample [25]. Future work

XSL•FO
RenderX

should evaluate SEE-Diabetes in larger and more diverse populations and test its implementation across various clinical settings and regions. Despite these limitations, our study supports the feasibility and potential value of integrating personalized education into routine care through tools like SEE-Diabetes. The structured, user-informed content provided by SEE-Diabetes may help improve patient-provider communication, support patient self-management, and ultimately contribute to more patient-centered chronic disease care.

Future Directions

Beyond the current implementation, SEE-Diabetes has potential for broader scalability across diverse care settings. While this study focused on EHR-based delivery, future work could explore parallel formats such as printable summaries or patient-facing handouts to support clinics without advanced EHR functionality, including rural and resource-limited programs. Additionally, situating SEE-Diabetes within national DSMES Standards and the 4 critical times for DSMES delivery may help align its use with formal education pathways while reinforcing self-management communication during routine care.

Conclusions

This study validated SEE-Diabetes, a patient-centered tool that embeds tailored diabetes self-management support into EHR notes for older adults. Both patients and clinicians confirmed that the content is clear, relevant, and feasible for integration into primary and specialty care. Embedding plain-language education within routine documentation may strengthen communication, reinforce self-care, and support chronic disease management in aging populations. Future work should evaluate implementation across diverse settings and its impact on clinical outcomes, engagement, and scalability.

Acknowledgments

The authors are grateful to the patients and clinicians at the University of Missouri Health Care who generously contributed their time and insights to this study. The authors also thank the interdisciplinary team members and research assistants who supported data collection and analysis.

Funding

This work was supported by the Agency for Healthcare Research and Quality [R21HS028032]; and the National Institute of Diabetes and Digestive and Kidney Disorders [P30DK092950].

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the confidential nature of the qualitative interview data but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MSK (lead), UK (equal), MD (supporting), and SAB (supporting)
Data curation: PN
Formal analysis: PN (lead) and MSK (supporting)
Funding acquisition: MSK
Investigation: PN (lead), SD (equal), UK (equal), MD (equal), MSK (supporting), and SAB (supporting)
Methodology: MSK (lead), PN (supporting), SD (supporting), UK (supporting), MD (supporting), and SAB (supporting)
Project administration: MSK (lead)
Resources: MSK
Supervision: MSK
Validation: MSK
Visualization: PN (lead), SD (supporting), and UK (supporting)
Writing – original draft: PN (lead)
Writing – review and editing: MSK (lead), SD (supporting), UK (supporting), MD (supporting), SAB (supporting), and EJS (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representative participant quotations illustrating perceptions of clarity, relevance, usability, and clinical usefulness of SEE-Diabetes content, organized by stakeholder group (patients and clinicians) and ADCES7-aligned domains.
[DOCX File, 18 KB - diabetes_v11i1e83448_app1.docx]

References

XSL•FO
RenderX

1. National diabetes statistics report. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2026-01-05]
2. Davis J, Fischl AH, Beck J, et al. 2022 National standards for diabetes self-management education and support. Diabetes Care 2022 Feb 1;45(2):484-494. [doi: 10.2337/dc21-2396] [Medline: 35050365]
3. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943. [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
4. He X, Li J, Wang B, et al. Diabetes self-management education reduces risk of all-cause mortality in type 2 diabetes patients: a systematic review and meta-analysis. Endocrine 2017 Mar;55(3):712-731. [doi: 10.1007/s12020-016-1168-2] [Medline: 27837440]
5. Alliston P, Jovkovic M, Khalid S, Fitzpatrick-Lewis D, Ali MU, Sherifali D. The effects of diabetes self-management programs on clinical and patient reported outcomes in older adults: a systematic review and meta-analysis. Front Clin Diabetes Healthc 2024;5:1348104. [doi: 10.3389/fcdhc.2024.1348104] [Medline: 38952998]
6. Narindrarangkura P, Boren SA, Khan U, Day M, Simoes EJ, Kim MS. SEE-diabetes, a patient-centered diabetes self-management education and support for older adults: findings and information needs from providers' perspectives. Diabetes Metab Syndr 2022 Sep;16(9):102582. [doi: 10.1016/j.dsx.2022.102582] [Medline: 35963033]
7. Narindrarangkura P, Boren SA, Khan U, Day M, Simoes EJ, Kim MS. SEE-Diabetes, a patient-centered diabetes self-management education and support for older adults: findings and information needs from patients' perspectives. Prim Care Diabetes 2022 Jun;16(3):395-403. [doi: 10.1016/j.pcd.2022.02.008] [Medline: 35227635]
8. ElSayed NA, McCoy RG, Aleppo G. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S86-S127. [doi: 10.2337/dc25-S005] [Medline: 39651983]
9. Graham AK, Munson SA, Reddy M, et al. Integrating user-centered design and behavioral science to design a mobile intervention for obesity and binge eating: mixed methods analysis. JMIR Form Res 2021 May 10;5(5):e23809. [doi: 10.2196/23809] [Medline: 33970114]
10. Freeman S, Lecka L, Davis K, et al. Leveraging user-centered design and usability and skills testing for a novel diabetes survival skills toolkit. J Endocr Soc 2025 Sep;9(9):bvaf111. [doi: 10.1210/jendso/bvaf111] [Medline: 40799775]
11. Ye Q, Patel R, Khan U, Boren SA, Kim MS. Evaluation of provider documentation patterns as a tool to deliver ongoing patient-centred diabetes education and support. Int J Clin Pract 2020 Mar;74(3):e13451. [doi: 10.1111/ijcp.13451] [Medline: 31769903]
12. Ghorayeb A, Darbyshire JL, Wronikowska MW, Watkinson PJ. Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach. BMJ Open 2023 Jan 30;13(1):e065323. [doi: 10.1136/bmjopen-2022-065323] [Medline: 36717136]
13. Facts/figures. MU Health Care. URL: https://www.muhealth.org/for-media/resources/factsfigures [accessed 2026-01-05]
14. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis: SAGE Publications; 2011. [doi: 10.4135/9781483384436]
15. American Association of Diabetes Educators. An effective model of diabetes care and education: revising the AADE7 self-care behaviors®. Diabetes Educ 2020 Apr;46(2):139-160. [doi: 10.1177/0145721719894903] [Medline: 31928334]
16. Patel D, Msosa YJ, Wang T, et al. An implementation framework and a feasibility evaluation of a clinical decision support system for diabetes management in secondary mental healthcare using CogStack. BMC Med Inform Decis Mak 2022 Apr 14;22(1):100. [doi: 10.1186/s12911-022-01842-5] [Medline: 35421974]
17. Wasylewicz A, Scheepers-Hoeks A. Clinical decision support systems. In: Dumontier M, Kubben P, Dekker A, editors. Fundamentals of Clinical Data Science: Springer, Cham; 2018:153-169. [doi: 10.1007/978-3-319-99713-1_11] [Medline: 31314237]
18. Shi Z, Du X, Li J, Hou R, Sun J, Marohabutr T. Factors influencing digital health literacy among older adults: a scoping review. Front Public Health 2024;12:1447747. [doi: 10.3389/fpubh.2024.1447747] [Medline: 39555039]
19. Johnson A. Health literacy, does it make a difference? Aust J Adv Nurs 2014;31(3):39-45. [doi: 10.37464/2016.334.1596]
20. Joyce MF, Berg S, Bittner EA. Practical strategies for increasing efficiency and effectiveness in critical care education. World J Crit Care Med 2017 Feb 4;6(1):1-12. [doi: 10.5492/wjccm.v6.i1.1] [Medline: 28224102]
21. Klein JW, Jackson SL, Bell SK, et al. Your patient is now reading your note: opportunities, problems, and prospects. Am J Med 2016 Oct;129(10):1018-1021. [doi: 10.1016/j.amjmed.2016.05.015] [Medline: 27288854]
22. Zhang A, Wang J, Wan X, et al. Community-based intelligent blood glucose management for older adults with type 2 diabetes based on the health belief model: randomized controlled trial. JMIR Mhealth Uhealth 2025 Jun 19;13:e60227. [doi: 10.2196/60227] [Medline: 40536934]
23. Stamate A, Marzan MD, Velciu M, Paul C, Spiru L. Advancing user-centric design and technology adoption for aging populations: a multifaceted approach. Front Public Health 2024;12:1469815. [doi: 10.3389/fpubh.2024.1469815] [Medline: 39712308]
24. Kim MS, Khan U, Boren SA, Narindrarangkura P, Ye Q, Simoes EJ. Transforming AADE7 for use in an evaluation framework for health information technology in diabetes mellitus. J Diabetes Sci Technol 2022 May;16(3):764-770. [doi: 10.1177/1932296820985842] [Medline: 33435720]

XSL•FO
RenderX

25. Census profile: midwest region. Census Reporter. 2021. URL: http://censusreporter.org/profiles/02000US2-midwest-region/ [accessed 2026-01-05]

Abbreviations

ADCES7: Association of Diabetes Care & Education Specialists 7 Self-Care Behaviors
DSMES: diabetes self-management education and support
EHR: electronic health record
SEE-Diabetes: Support-Engage-Empower-Diabetes
UCD: user-centered design

Edited by S Li; submitted 03.Sep.2025; peer-reviewed by AH Fischl, AD Keetharuth; revised version received 06.Jan.2026; accepted 07.Jan.2026; published 06.Feb.2026.

Please cite as:
Narindrarangkura P, Dejhansathit S, Khan U, Day M, Boren SA, Simoes EJ, Kim MS
Content Validation of an Electronic Health Record–Based Diabetes Self-Management Support Tool for Older Adults With Type 2 Diabetes: Qualitative Study
JMIR Diabetes 2026;11:e83448
URL: https://diabetes.jmir.org/2026/1/e83448
doi:10.2196/83448

© Ploypun Narindrarangkura, Siroj Dejhansathit, Uzma Khan, Margaret Day, Suzanne A Boren, Eduardo J Simoes, Min Soon Kim. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 6.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cardiovascular-Kidney-Metabolic Syndrome: Development of an ICD-10-CM Coding Framework

Minzhe Zhao[1*], MSc; Junhao Zhang[1*], MM; Huiyun Wang[2], MM; Chongjun Xu[1], MM; Jiahui Li[1], MM; Chenghua Li[3], MSN; Xuemei He[4], MSN; Xueying Zheng[1], MD

[1]Department of Endocrinology and Metabolism, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, 17 Lujiang Road, Luyang District, Hefei, Anhui, China

[2]Graduate School, Bengbu Medical College, Bengbu, Anhui, China

[3]Pan-Vascular Management Center, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China

[4]Outpatient Department, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China

[*]these authors contributed equally

Corresponding Author:
Xueying Zheng, MD
Department of Endocrinology and Metabolism, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, 17 Lujiang Road, Luyang District, Hefei, Anhui, China

Abstract

Background: Cardiovascular-kidney-metabolic (CKM) syndrome is a multisystem construct describing the intertwined progression of cardiometabolic risk factors, chronic kidney disease (CKD), and cardiovascular disease. The American Heart Association (AHA) proposed CKM stages (0‐4) for risk stratification and prevention. However, CKM lacks a single *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) code, which hinders standardized stage identification in electronic health records and claims data.

Objective: This study aimed to develop an AHA-aligned *ICD-10-CM* coding framework as an implementation template that operationalizes CKM stages 0‐4 for reproducible cohort identification and stage-based analyses in real-world data.

Methods: We mapped AHA CKM stages (0‐4) to *ICD-10-CM* diagnosis code sets using Fiscal Year 2026 conventions, code-set engineering best practices, and clinician review. To improve reproducibility, we defined a hierarchical staging algorithm, co-occurrence rules, and recommended lookback and encounter-confirmation thresholds. Stage 3 includes guidance for electronic health record–enhanced ascertainment and claims-only proxies.

Results: We provide stage-specific *ICD-10-CM* code sets for CKM stages 0‐4. Stage 1 captures excess or dysfunctional adiposity or prediabetes. Stage 2 captures established metabolic disease and earlier-stage CKD. Stage 3 captures subclinical cardiovascular injury or very high-risk CKD. Stage 4 captures overt clinical cardiovascular disease events, with or without kidney failure.

Conclusions: This implementation framework enables transparent, reproducible CKM staging in real-world datasets and supports stage-based epidemiologic and health-system applications. Empirical validation and local implementation testing are needed before clinical deployment.

(JMIR Diabetes 2026;11:e91827) doi:10.2196/91827

KEYWORDS

cardiovascular-kidney-metabolic syndrome; CKM syndrome; ICD-10-CM; coding; risk stratification; chronic kidney disease; cardiovascular disease; metabolic syndrome; diabetes mellitus; hypertension

Introduction

The Cardiovascular-Kidney-Metabolic Syndrome: An Integrative Health Construct

Cardiovascular-kidney-metabolic (CKM) health describes bidirectional interactions among metabolic risk factors, chronic kidney disease (CKD), and cardiovascular disease (CVD) that can mutually amplify disease progression [1,2]. CKM health is not a separate syndrome; it reframes the common clustering of obesity, type 2 diabetes mellitus (T2DM), and hypertension that accelerates both renal and cardiac decline [3-5]. CKM often begins with excess or dysfunctional adiposity and is characterized by the clustering of dysglycemia, hypertension, dyslipidemia, and systemic inflammation. These processes jointly accelerate damage to the heart, kidney, and vasculature

XSL•FO
RenderX

and increase the risk of adverse cardio-renal events [6]. The American Heart Association (AHA) has therefore emphasized a shift from siloed management to stage-based prevention across the CKM continuum.

CKM syndrome reflects a reinforcing cycle—excess or dysfunctional adiposity, insulin resistance, and chronic inflammation. These factors create hemodynamic and metabolic stress on the kidneys and vasculature. Adipose tissue–derived inflammatory mediators and oxidative stress promote hypertension, dyslipidemia, and T2DM, which in turn accelerate CKD progression and atherosclerotic and heart failure (HF) phenotypes. This integrated construct motivates stage-based prevention and treatment strategies that address upstream drivers and downstream target-organ injury rather than treating each condition in isolation [3].

The AHA Staging Framework for CKM (Stages 0-4)

The AHA proposed a 5-stage CKM framework (stages 0‑4) that links upstream cardiometabolic risk to downstream subclinical and clinical cardiovascular and kidney disease [1,2]. These stage definitions provide a shared language for risk stratification and support stage-appropriate prevention and management across the CKM continuum [3].

The AHA CKM staging scheme is defined as follows [1,2]:

- Stage 0: No CKM risk factors. Individuals with normal BMI and waist circumference, normoglycemia, normotension, a normal lipid profile, and no evidence of CKD or subclinical or clinical CVD.
- Stage 1: Excess or dysfunctional adiposity. Individuals with overweight or obesity, abdominal obesity, or dysfunctional adipose tissue, without the presence of other metabolic risk factors or CKD. This includes BMI≥25 kg/m² (or ≥23 kg/m² if Asian ancestry), waist circumference ≥88 cm in women or ≥102 cm in men (or ≥80 cm in women and ≥90 cm in men if Asian ancestry), or dysfunctional adiposity defined as hyperglycemia or prediabetes (fasting blood glucose 100‑124 mg/dL or hemoglobin A1c 5.7%‑6.4%).
- Stage 2: Metabolic risk factors and CKD. Individuals with metabolic risk factors (hypertriglyceridemia, hypertension, metabolic syndrome, and diabetes), or moderate- to high-risk CKD (per Kidney Disease: Improving Global Outcomes [KDIGO] eGFR [estimated glomerular filtration rate]-albuminuria risk categories).
- Stage 3: Subclinical CVD in CKM. Subclinical atherosclerotic cardiovascular disease (ASCVD) or subclinical HF among individuals with excess or dysfunctional adiposity, other metabolic risk factors, or CKD. Subclinical ASCVD is principally diagnosed by coronary artery calcification (subclinical atherosclerosis by coronary catheterization or computed tomography [CT] angiography also meets criteria). Subclinical HF can be diagnosed by elevated cardiac biomarkers (NT-proBNP ≥125 pg/mL; hs-troponin *t*≥14 ng/L for women and ≥22 ng/L for men; hs-troponin I≥10 ng/L for women and ≥12 ng/L for men) and/or echocardiographic parameters; a combination of biomarker and echocardiographic abnormalities indicates the highest HF risk. Risk equivalents include very high-risk CKD (G4-G5 CKD or very high risk per KDIGO) and high predicted 10-year CVD risk.
- Stage 4: Clinical CVD in CKM. Clinical CVD (coronary heart disease, HF, stroke, peripheral artery disease (PAD), and atrial fibrillation) among individuals with excess or dysfunctional adiposity, other CKM risk factors, or CKD. Stage 4 is further divided into 4A (no kidney failure) and 4B (kidney failure present).

Rationale for an *ICD-10* or *ICD-10-CM* Coding Approach

A key barrier to implementing CKM in practice is the absence of a single *ICD-10* (*International Classification of Diseases, Tenth Revision*) code that represents the entire syndrome. This limits standardized CKM identification in electronic health record (EHR) and claims-based systems used for clinical care, population health management, and epidemiologic research [7-10]. Without a CKM-specific *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) code, case identification is inconsistent, limiting prevalence estimation, treatment evaluation, and large-scale studies of CKM natural history and outcomes [11,12]. As a result, there is a need for developing clear, reproducible, and AHA-compliant operational definitions to create mapping between *ICD-10-CM* codes and CKM's various pathological stages. This will support more consistent identification of CKM-related phenotypes, more transparent stage assignment, and more reproducible cohort construction and outcome ascertainment across health care systems and research datasets [1,2].

There are several reasons for establishing a uniform *ICD-10-CM* approach to operationalize CKM stages. First, standardized coding enables automated identification of patients who meet stage 1‑2 criteria (eg, excess or dysfunctional adiposity, prediabetes, or hypertension) for screening, prevention programs, and longitudinal monitoring in routine care settings [13]. Second, consistent code sets support reproducible cohort construction and outcome ascertainment across health systems, improving comparability of observational studies and enabling pragmatic evaluation of prevention and treatment strategies in real-world data environments [14-16]. Finally, a shared staging framework facilitates communication across multidisciplinary teams and aligns documentation, analytics, and care pathways around the AHA CKM construct [15].

Study Objectives

The objectives of this manuscript are to (1) present an AHA-aligned *ICD-10-CM* code set mapping CKM stages 0‑4 to existing diagnosis codes, (2) specify operational rules and assumptions for applying the algorithm in claims and EHR data, and (3) summarize implementation considerations and limitations to support reproducible use and future validation.

Methods

Code Set Development and Sources

We aligned CKM stage definitions to the 2023 AHA CKM construct and translated stage-defining phenotypes and outcomes into *ICD-10-CM* diagnosis code sets. Code candidates were

XSL•FO
RenderX

assembled from the *ICD-10-CM* Tabular List and Index and refined using code-set engineering guidance and clinical domain knowledge. The framework is explicitly scoped to US *ICD-10-CM* (not ICD-10-CN or other national modifications). We used Fiscal Year 2026 (FY2026) *ICD-10-CM* conventions (effective October 1, 2025) to reflect current code structures and recent additions (eg, adult obesity class codes and T2DM in remission). For implementation and reproducibility, the full stage-specific code sets and logic specifications are provided in machine-actionable appendix files, with JSON as the authoritative specification and CSV as a flattened companion file for review and audit (Multimedia Appendix 1).

Algorithm Rules and Operational Assumptions

To improve reproducibility, we specify a hierarchical staging rule in which the highest qualifying stage is assigned (stage 4 supersedes 3, which supersedes 2, etc). For claims-based implementations, chronic conditions (eg, obesity, hypertension, and CKD) can be defined using ≥2 outpatient claims on different dates (eg, ≥30 d apart) or ≥1 inpatient claim within a prespecified lookback window (eg, 1‑2 y), whereas acute events (eg, myocardial infarction and stroke) may be defined using a single inpatient claim. Stage eligibility is determined by stage-specific co-occurrence rules aligned with AHA definitions. Stage 1 requires excess or dysfunctional adiposity without other metabolic risk factors or CKD. Stage 2 requires metabolic risk factors and/or moderate- to high-risk CKD. Stage 3 requires subclinical ASCVD or subclinical HF—or risk equivalents (very high-risk CKD [G4–G5] or high predicted 10-year CVD risk)—in individuals with adiposity, other metabolic risk factors, or CKD. Stage 4 requires clinical CVD co-occurring with adiposity, other CKM risk factors, or CKD, with optional 4A or 4B stratification by kidney failure. For claims-only implementations, KDIGO risk strata are approximated using *ICD-10-CM* proxies (primarily N18 stage codes, with R80 as an optional proxy for albuminuria or proteinuria when documented); when structured laboratory data are available, eGFR and albumin-to-creatinine ratio can be used to compute KDIGO eGFR-albuminuria risk categories directly. Because subclinical disease is incompletely captured in claims data, stage 3 should be implemented preferentially in an EHR-enhanced framework, or in EHR-only settings where structured laboratory, imaging, or echocardiographic data are available. Claims-only stage 3 should be treated as a conservative proxy, primarily reflecting very high-risk CKD and related risk-equivalent states, and should be labeled as such in analytic applications (Table 1). Stage 0 is defined clinically by normal anthropometrics and cardiometabolic profiles and is operationalized in claims as the absence of qualifying codes for stages 1‑4 within the lookback window. We followed the FY2026 *ICD-10-CM* Official Guidelines for Coding and Reporting for code selection and convention handling [17].

XSL•FO
RenderX

Table . Summary of CKM[a] stages and corresponding *ICD-10-CM*[b] codes.[c,d,e]

CKM stage and domain or category	Operational definition and *ICD-10-CM* code set
Stage 0	
Clinical definition (AHA[f])	Individuals with normal BMI and waist circumference, normoglycemia, normotension, a normal lipid profile, and no evidence of CKD[g] or subclinical or clinical CVD[h].
Claims-based operationalization (*ICD-10-CM*)	No stage-defining diagnosis codes from stages 1‑4 are present within the lookback window.
Stage 1	
Clinical definition (AHA)	Individuals with overweight or obesity, abdominal obesity, or dysfunctional adipose tissue, without the presence of other metabolic risk factors or CKD. Criteria include BMI ≥25 kg/m^2 (or ≥23 kg/m^2 if Asian ancestry), waist circumference ≥88 cm in women and ≥102 cm in men (or if Asian ancestry ≥80 cm and ≥90 cm in women and men, respectively), or dysfunctional adiposity defined as hyperglycemia or prediabetes (fasting blood glucose 100‑124 mg/dL or HbA1c[i] 5.7%‑6.4%).
Claims-based operationalization (*ICD-10-CM*)	Overweight or obesity: overweight (E66.3), obesity class 1‑3 (E66.811-E66.813), and other obesity codes (E66.8, E66.9); adult BMI codes (Z68.25-Z68.45) may be used as optional supporting documentation when available; dysfunctional adiposity proxy: prediabetes (R73.03).
Exclusion rule	Absence of stage 2‑4 core codes in the lookback window.
Supplementary (non–stage-defining) adiposity-related codes (optional)	Abnormal weight gain (R63.5); other insulin resistance codes if used locally.
Stage 2	
Clinical definition (AHA)	Individuals with metabolic risk factors (hypertriglyceridemia [≥135 mg/dL], hypertension, metabolic syndrome, and diabetes), or moderate- to high-risk CKD.
Hypertension	Essential hypertension (I10), hypertensive heart disease without heart failure (I11.9), hypertensive CKD (I12.-+ N18.-), hypertensive heart and CKD without heart failure (I13.10).
Hypertriglyceridemia	Pure hypertriglyceridemia (E78.1); expanded claims-based proxy may additionally include mixed hyperlipidemia (E78.2) to capture common mixed triglyceride-elevated phenotypes in administrative data.
Metabolic syndrome	E88.810.
Diabetes	E11.- (including uncomplicated and remission code E11.A) and other diabetes codes as appropriate (eg, E10.-, E13.-).
CKD (moderate- to high-risk; claims proxy)	CKD stage 3 (N18.30-N18.32) or CKD stage 1‑2 (N18.1-N18.2) with albuminuria and proteinuria codes (eg, R80.-) within the lookback window; exclude very high-risk CKD (N18.4-N18.5) and kidney failure (N18.6).
Supplementary risk-enhancing conditions (not stage-defining)	Nonalcoholic fatty liver disease or NASH[j] (K76.0, K75.81); sleep apnea disorders (G47.30, G47.31, G47.33, G47.39); polycystic ovary syndrome (E28.2).
Stage 3	
Clinical definition (AHA)	Subclinical ASCVD[k] or subclinical HF[l] among individuals with excess or dysfunctional adiposity, other metabolic risk factors, or CKD. Subclinical ASCVD is principally diagnosed by coronary artery calcification (subclinical atherosclerosis by coronary catheterization or CT[m] angiography also meets criteria). Subclinical HF is diagnosed by elevated cardiac biomarkers and/or echocardiographic parameters, with a combination of the 2 indicating highest HF risk. Risk equivalents include very high-risk CKD (G4-G5 CKD or very high risk per KDIGO[n]) and high predicted 10-year CVD risk.
Claims-based operationalization (*ICD-10-CM*)	Claims-based operationalization (*ICD-10-CM* core risk equivalents): very high-risk CKD (N18.4-N18.5). Claims-only proxies for subclinical ASCVD or HF are limited and may under-ascertain stage 3.

XSL•FO
RenderX

CKM stage and domain or category	Operational definition and *ICD-10-CM* code set
Supplementary CKD-related codes (optional)	CKD-related secondary hyperparathyroidism of renal origin (N25.81) when co-occurring with N18.4-N18.5.
Stage 4	
Clinical definition (AHA)	Clinical CVD (coronary heart disease, HF, stroke, peripheral artery disease, atrial fibrillation) among individuals with excess or dysfunctional adiposity, other CKM risk factors, or CKD; stage 4A: no kidney failure; stage 4B: kidney failure present.
Claims-based operationalization (*ICD-10-CM*)	Assign stage 4 only when a clinical CVD code co-occurs with any stage 1‐3 core code and/or CKD within the lookback window.
Coronary heart disease or angina	Chronic atherosclerotic coronary disease (I25.-), angina pectoris (I20.-).
AMI[o]	AMI (I21.-; includes I21.0–I21.3 [STEMI[p]], I21.4 [NSTEMI[q]], I21.A1 [type 2 MI[r]]), subsequent MI (I22.-).
HF	Left ventricular failure (I50.1), systolic (congestive) HF (I50.2-), diastolic HF/HFpEF (I50.3-), combined systolic & diastolic HF (I50.4-), HF unspecified (I50.9); hypertensive heart disease with HF (I11.0+ I50.-); hypertensive heart and CKD phenotype with HF (I13.0/I13.2+ I50.-+ N18.-).
AF[s]	Paroxysmal AF (I48.0), persistent AF (I48.19), chronic AF (I48.20), permanent AF (I48.21).
Peripheral artery disease	Atherosclerosis of native arteries of extremities (I70.2-), peripheral vascular disease unspecified (I73.9).
Stroke (acute events)	Nontraumatic subarachnoid hemorrhage (I60.-), nontraumatic intracerebral hemorrhage (I61.-), other and unspecified nontraumatic intracranial hemorrhage (I62.-), cerebral infarction/ischemic stroke (I63.-), stroke not specified as hemorrhage or infarction (I64), sequelae of cerebrovascular disease (I69.-), personal history of TIA[t] and cerebral infarction without residual deficits (Z86.73).
Claims-based operationalization (*ICD-10-CM*)	End-stage renal disease (N18.6), dependence on renal dialysis (Z99.2). Kidney transplant status (Z94.0+ N18.-).

XSL•FO
RenderX

CKM stage and domain or category	Operational definition and *ICD-10-CM* code set
Kidney failure (stage 4b)	ESRD[u] (N18.6) and/or dialysis dependence (Z99.2), with optional transplant status codes when relevant.

[a]CKM: cardiovascular-kidney-metabolic.

[b]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification*.

[c]Notation: In this manuscript, a hyphen suffix denotes a code-family prefix rather than an automatically executable rule. Whether descendants are included, and whether exclusions apply, is specified explicitly in the accompanying machine-actionable appendix files (JSON/CSV). Implementers should not infer descendant inclusion from shorthand notation alone.

[d]The definition of metabolic syndrome (MetS) is the presence of ≥3 of the following: (1) waist circumference ≥88 cm for women and ≥102 cm for men (≥80 cm for women and ≥90 cm for men if Asian ancestry); (2) HDL-C <40 mg/dL in men or <50 mg/dL in women; (3) triglycerides ≥150 mg/dL; (4) elevated blood pressure (systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥80 mmHg and/or antihypertensive medication use); (5) fasting glucose ≥100 mg/dL.

[e]Operational assumptions (recommended for reproducibility): assign the highest qualifying stage (4>3>2>1>0). For chronic conditions, require ≥2 outpatient claims on different dates or ≥1 inpatient claim within a defined lookback window; for acute events, a single inpatient claim may be sufficient. Implementations should prespecify timing windows and setting rules (inpatient vs outpatient) based on the data source. For Stage 3, EHR-enhanced ascertainment is preferred whenever structured laboratory, imaging, or echocardiographic data are available.

[f]AHA: American Heart Association.

[g]CKD: chronic kidney disease.

[h]CVD: cardiovascular disease.

[i]HbA_{1c}: hemoglobin A_{1c}.

[j]NASH: nonalcoholic steatohepatitis.

[k]ASCVD: atherosclerotic cardiovascular disease.

[l]HF: heart failure.

[m]CT: computed tomography.

[n]KDIGO: Kidney Disease: Improving Global Outcomes.

[o]AMI: acute myocardial infarction.

[p]STEMI: ST-segment elevation myocardial infarction.

[q]NSTEMI: non–ST-segment elevation myocardial infarction.

[r]MI: myocardial infarction.

[s]AF: atrial fibrillation.

[t]TIA: transient ischemic attack.

[u]ESRD: end-stage renal disease.

Validation and Quality Assurance

We conducted internal code verification to remove non-*ICD* (*International Classification of Diseases*) labels, correct formatting, and ensure consistency with current *ICD-10-CM* conventions. We also compared the stage logic and key code families with published CKM implementation studies and relevant guideline definitions. This manuscript presents an implementation framework supported by literature review and expert verification. Empirical evaluation against chart review, registry linkage, or structured EHR reference standards remains an important next step for future work.

For claims-only implementations, we recommend defining an index date (eg, cohort entry) and applying a lookback window (eg, 12 - 24 mo) to capture qualifying diagnoses. A common reproducibility pattern is to require ≥1 inpatient claim or ≥2 outpatient claims on separate dates for chronic conditions, such as hypertension, diabetes, and CKD, and then assign stage using a hierarchical rule (stage 4>stage 3>stage 2>stage 1>stage 0) so that downstream clinical events supersede upstream risk states. Sensitivity analyses can vary the lookback window and diagnosis-claim thresholds to evaluate robustness.

For EHR-enhanced implementations (preferred for stage 3), structured laboratory values and imaging results can be integrated to capture subclinical disease definitions (eg, elevated natriuretic peptides, abnormal troponin, and coronary artery calcium). When these data are unavailable, stage 3 should be treated as a conservative proxy based on available diagnosis codes, and this limitation should be explicitly reported in studies using claims-only data.

Below we illustrate a claims-based use case to demonstrate how the staging hierarchy, lookback window, and encounter-confirmation rules translate into a final CKM stage assignment.

Results

Overview

Table 1 presents the AHA stage definitions (0 - 4) alongside claims-based *ICD-10-CM* operationalizations, with risk-enhancing conditions listed separately as supplementary (non–stage-defining) codes.

Implementation and Use Cases

For example (claims-only), consider a 60-year-old patient with an index date of January 1, 2026, and a 24-month lookback window. Chronic conditions are confirmed by ≥2 outpatient

XSL•FO
RenderX

claims on different dates (≥30 d apart) or ≥1 inpatient claim; acute events are confirmed by a single inpatient claim.

Within the lookback window, the patient has (1) two outpatient claims for essential hypertension (I10) 45 days apart, (2) two outpatient claims for T2DM without complications (E11.9) 60 days apart, (3) one outpatient claim for obesity class 3 (E66.813) with a BMI code (Z68.41), and (4) one inpatient claim for acute myocardial infarction (AMI; I21.3). No codes for kidney failure (N18.6, Z99.2) are present.

Stage assignment: The confirmed I10 or E11.9 (and obesity) qualify the patient for stage 2 (established metabolic disease). The inpatient AMI code (I21.3) is a clinical CVD event and therefore meets stage 4 criteria. Applying the hierarchical rule (4>3>2>1>0), the patient is assigned CKM stage 4A (clinical CVD without kidney failure).

For contrast, a patient with overweight or obesity (E66.3, E66.8, E66.9, and/or BMI Z68.25-Z68.29) or prediabetes (R73.03) and no stage 2‑4 core codes in the same lookback window would be assigned stage 1.

CKM Staging and Core Metabolic Disorders: *ICD-10-CM* Coding

CKM syndrome is a multidomain construct that spans metabolic risk factors, kidney disease, and CVD [18-22]. In this section, we summarize the major *ICD-10-CM* coding domains relevant to CKM staging and risk stratification. Some conditions are stage-defining within the proposed framework, whereas others serve as risk-enhancing or supplementary features that may refine phenotyping and implementation in real-world data.

Hypertension

Hypertension represents one of the most common and important cardiovascular risk factors in CKM syndrome [23]. As one of the key elements of CKM syndrome, hypertension is a driver of the long-term and poorly controlled coronary artery disease, myocardial infarction, stroke, and peripheral arterial disease that characterize this condition [24]. The long-term elevation of blood pressure (BP) causes an excessive workload on the heart and results in an abnormal increase in heart size (hypertrophy) and ultimately acute HF [25,26]. Also, an elevated BP is a primary factor in the development of atherosclerosis, ultimately leading to increased risk for coronary arterial disease, stroke, and peripheral arterial disease [27,28]. As a result, hypertension serves as a double-edged sword to both create CKD and worsen the control of BP due to the resulting kidney damage [29]. With the availability of the *ICD-10-CM* codes, which are detailed and hierarchical for hypertension, it is possible to document the presence of hypertension and its associated complications accurately and appropriately. Additionally, knowing the specific codes that are used in conjunction with CKM staging will be the basis for accurate clinical assessments and appropriate management of all patients with CKM syndrome [1].

In *ICD-10-CM*, hypertensive heart disease is coded under I11 (I11.0 for hypertensive heart disease with HF, I11.9 without HF). When hypertension coexists with CKD, hypertensive CKD is coded under I12 with an accompanying N18 code to specify CKD stage. Combined hypertensive heart and CKD phenotypes are captured under I13 and should be paired with I50 (HF) and N18 (CKD stage) as applicable [17]. These conventions (rather than non-*ICD* labels) improve consistency when operationalizing CKM stage 2‑4 phenotypes in EHR and claims data.

Essential Hypertension (I10)

The *ICD-10-CM* code I10 is for essential hypertension [17]. This means that a person with a diagnosis of high BP does not have any evidence of targets (such as CVD or CKD). As per CKM and CKD staging framework, a diagnosis of essential hypertension (without other metabolic risk factors or CKD) puts that person in CKM stage 2. This is due to the fact that essential hypertension is a well-established metabolic risk factor, which greatly increases the chance of developing both CVD and CKD [24,29]. Coding patients with the specific I10 code can help identify these individuals as early CKM stage 2 patients who may benefit from lifestyle modification and the use of an antihypertensive to prevent progress to later CKM stages [30]. The code I10 should only be assigned if a qualified health care provider has formally diagnosed hypertension and does not rely upon a single elevated BP measurement only [18,31]. Additionally, documentation must include evidence that the patient has persistently elevated BP that meets the criteria for a diagnosis of hypertension.

The clinical importance of identifying and coding for essential hypertension cannot be underestimated. In fact, it typically represents the first manifestation of a progressive CKM syndrome and is therefore an opportune time for health care providers to initiate treatment. It is also essential to enabling health care providers to monitor and treat patients at risk of other metabolic risk factors, such as dyslipidemia and T2DM, as well as identify early signs of target-organ injury. Early identification and appropriate intervention are critical to avoid the long-term complications associated with hypertension, which are often serious and debilitating [30]. In addition, the use of a specific code for essential hypertension will facilitate tracking the prevalence of this disease and evaluating the success of public health efforts aimed at decreasing the prevalence of hypertension in the population [32]. With the use of EHR systems, code I10 may trigger clinical decision support tools to provide reminders for BP testing, lifestyle modifications, and medication management, thus improving patient care for this major and common health problem [33,34].

Hypertensive Heart Disease (I11)

Hypertensive heart disease is identified by *ICD-10-CM* category I11 when the medical record indicates that the patient's chronic high BP has caused some level of structural or functional impairment to their heart [17]. Hypertensive heart disease represents a more advanced stage of disease than essential hypertension and indicates that the patient's heart has already incurred damage due to being subjected to high BP for a prolonged period of time [35]. The most common manifestation of hypertensive heart disease is left ventricular hypertrophy (LVH), which indicates that the heart muscle has become thicker in response to increased workload. Other possible manifestations of hypertensive heart disease are HF, a condition where the heart can no longer pump sufficient amounts of blood to supply

XSL•FO
RenderX

the body's requirements [36], and would also require coding from the I50 category to specify the type of HF that is present (eg, left-sided HF [I50.1] or HF with preserved ejection fraction [I50.3]) [37]. Hypertensive heart disease should be staged according to clinical phenotype rather than diagnosis label alone. Hypertension without evidence of cardiac structural or functional involvement is consistent with CKM stage 2. Hypertensive heart disease without HF, when it reflects subclinical cardiac remodeling or dysfunction, is more appropriately aligned with CKM stage 3. Hypertensive heart disease with HF should be classified as CKM stage 4 [1,2].

The I11 code category is essential in accurately capturing the patient's level of clinical severity and in determining how to best manage their care. Patients with hypertensive heart disease are at a substantially increased risk for adverse cardiovascular events (such as myocardial infarction, stroke, and sudden cardiac death) compared with patients with essential hypertension alone [36]. Therefore, patients with hypertensive heart disease require more intensive monitoring and treatment, including aggressive management of BP and lifestyle modification, and often including treatment for HF or other cardiac complications [38]. To support the application of the code for hypertensive heart disease, the provision of satisfactory documentation in the patient's medical record must connect the patient's recorded hypertension with their associated cardiac findings. Documentation must include any diagnostic tests performed to evaluate the patient's condition (an echocardiogram may be used to identify LVH) or clinical findings that correlate with HF [39]. By properly coding for hypertensive heart disease, health care providers can offer their patients appropriate care levels and monitor their treatment progress over time [40]. Additionally, the ability to code for hypertensive heart disease will aid researchers in identifying populations at greater risk and in developing methods to treat patients diagnosed with this formidable condition.

Hypertensive CKD (I12)

The *ICD-10-CM* code category I12 indicates hypertension with CKD, helping distinguish hypertensive CKD from other etiologies (eg, diabetes, glomerulonephritis, and polycystic kidney disease) [17,20]. Per *ICD-10-CM* conventions, an additional N18 code should be assigned to specify the CKD stage based on clinician-documented CKD stage. In our CKM operationalization, I12 supports the presence of CKD, and the resulting CKM stage depends on CKD severity (eg, N18.1-N18.3 aligns with CKM stage 2, N18.4-N18.5 with stage 3) and whether clinical CVD codes are present (stage 4). Thus, hypertensive CKD reflects hypertensive target-organ involvement, but CKM stage assignment should follow the predefined hierarchical staging rules rather than treating I12 alone as stage 3 or 4.

As for the clinical impact of hypertensive CKD, there is a strong reciprocal relationship between hypertension and CKD. Patients with hypertensive CKD may experience accelerated progression of CKD due to lack of adequate BP control as a result of declining kidney function, and the classes of medications for the management of hypertension and CKD have overlapping functions. Hypertensive CKD reflects a bidirectional hypertension-CKD interaction that accelerates renal decline and increases risks of CVD and end-stage renal disease (ESRD) [29,41]. Management is multifaceted and typically includes BP control (often renin-angiotensin-aldosterone system inhibition) and control of coexisting cardiometabolic risk factors [42]. To support I12 assignment, documentation should explicitly link hypertension as the etiology of CKD and note the rationale when alternative causes are unlikely [43,44].

Hypertensive Heart and CKD (I13)

ICD-10-CM category I13 captures combined hypertensive heart disease and CKD attributable to hypertension [17]. This phenotype indicates advanced cardio-renal target-organ injury, in which cardiac dysfunction can worsen renal perfusion and CKD can exacerbate HF via volume and electrolyte disturbances [45,46]. When using I13, documentation should specify HF type (I50) and CKD stage (N18) to preserve clinical granularity.

Patients with hypertensive heart disease and CKD have a complex medical management, because these patients typically take several medications, including multiple antihypertensives, diuretics for fluid overload, and medications for managing other cardiovascular risk factors. In addition, these patients require continuous monitoring of their fluid status, electrolyte levels, and generally, renal function. Treatment options are drastically reduced for patients with both cardiac and renal dysfunction, which leads to the need for careful risk-versus-benefit analysis for treatment options [45,47]. For instance, when considering the use of renin-angiotensin-aldosterone system inhibitors, which are of benefit to both the heart and kidney, the risk of hyperkalemia and/or further decline in kidney function may limit their usage [48]. Providers need to ensure that the medical record documents an established link between hypertension, heart disease, and CKD to support the use of the I13 code. Providers usually determine this connection through a combination of patient history, examination, and diagnostic test results. Correctly coding this complicated condition is necessary so that patients can receive the proper specialized care as well as to demonstrate to the public and for reimbursement the seriousness of their condition.

T2DM

The CKM syndrome model has identified T2DM as the prototype metabolic disorder. Thus, T2DM represents a chronic condition caused by both insufficient insulin and insulin resistance, which leads to hyperglycemia [49]. T2DM usually occurs along with all other important aspects of CKM syndrome, especially dyslipidemia, hypertension, and increased body fat; these combinations of the various risk factors create an extremely powerful synergy and thus significantly contribute to the rapid progression of both CKD and ASCVD [1,2]. Additionally, because of the continued presence of hyperglycemia resulting from T2DM, there is a direct harm to the kidneys, resulting in the development of diabetic nephropathy and a significant number of the individuals affected by this condition will go on to develop CKD and ESRD. Moreover, T2DM further increases endothelium dysfunction, promotes inflammation, and increases oxidative stress, all of which lead to the development of atherosclerosis [50]. Therefore, T2DM will greatly influence the future development

XSL•FO
RenderX

of myocardial infarction, stroke, and congestive HF seen in patients with CKM [51]. Thus, it can be concluded that the presence of T2DM without other complications would constitute an individual falling into the CKM stage 2 risk category.

ICD-10-CM provides a standardized diagnosis coding system for T2DM and its complications [17,52]. Using *ICD-10-CM* enables hierarchical documentation of T2DM and associated complications, providing granularity that supports operationalizing CKM staging in administrative data. Diabetes complications, such as nephropathy, are treated as advanced disease manifestations; thus, individuals with these diagnoses align with higher-risk CKM profiles. E11 is the base code category for T2DM; the fourth, fifth, and sixth characters specify the presence and type of complication, allowing more precise phenotyping. This coding granularity supports improved characterization of a patient's overall CKM health and evaluation of interventions aimed at preventing or treating diabetic complications [53].

T2DM Without Complications (E11.9, E11.A)

A patient diagnosed with T2DM may be assigned *ICD-10-CM* code E11.9 ("type 2 diabetes mellitus without complications") when there is no documented long-term complication such as nephropathy, retinopathy, or neuropathy [17,52]. Individuals with uncomplicated T2DM (E11.9) plus concurrent risk factors (eg, obesity and hypertension) typically meet criteria for CKM stage 2 (established metabolic disease). CKM stage 2 represents a key prevention and intervention window for intensive risk-factor modification and cardiometabolic protection [54]. In FY2026, *ICD-10-CM* introduced E11.A ("type 2 diabetes mellitus without complications, in remission"), which aligns conceptually with consensus definitions of T2DM remission and provides a standardized way to document remission status when applicable [17,55]. Accurate coding of uncomplicated T2DM and remission status can support stratified prevention strategies and improve reproducibility in real-world CKM research [56].

CKD

CKD is a key component of the CKM syndrome, being both a primary target organ lesion caused by metabolic dysfunction as well as a risk factor for CVD [1,2,57]. Although staging CKD is emphasized by both AHA and international guidelines using eGFR and albuminuria, follow-up appointments with nephrologists should still be performed at all stages of CKM [20]. The *ICD-10-CM* provides a set of codes for CKD that are critical for accurate risk stratification and management of patients with CKD, as well as all patients under CKM. The following codes represent the respective stages of CKD: N18.1 through N18.5 indicate CKD stages 1 through 5, while N18.6 indicates ESRD. Z94.0 denotes a patient having undergone a renal transplant [17]. Accurate recording of CKD codes is essential for following the patient's disease course and identifying patients with CKD at high risk of cardiovascular events so that timely and appropriate medical care can be provided by nephrologists. By using the most specific coding standards available, practitioners can better capture key CKD features in the EHR (eg, CKD stage [N18], kidney failure or dialysis dependence [N18.6 and Z99.2], transplant status [Z94.0], and albuminuria or proteinuria when documented [R80]). This supports individual care and population-based CKD research [58].

CKD Staging (N18.1-N18.5)

ICD-10-CM category N18 captures clinician-documented CKD stage (1-5) based on a patient's eGFR and the degree of albuminuria [17,20]. Albuminuria should be coded separately when documented (eg, R80). Stage 1 CKD (N18.1) represents early CKD, whereas stage 5 (N18.5) indicates advanced CKD immediately preceding kidney failure (N18.6). For clinical documentation, the assigned N18 code should match the clinician-documented CKD stage at the time of the encounter. For research applications in longitudinal data, investigators should prespecify whether CKD severity is defined using the current stage, a worst-ever stage within a lookback window, or time-updated staging, and report this choice explicitly. In our CKM operationalization, CKD stages 1‑3 generally align with CKM stages 2‑3, whereas CKD stages 4‑5 indicate very high-risk CKD and typically align with CKM stage 3 (or stage 4 when clinical CVD co-occurs) [44].

ESRD and Transplant Status (N18.6, Z99.2, and Z94.0)

ICD-10-CM code N18.6 refers to ESRD, the final stage of CKD requiring renal replacement therapy, such as dialysis or transplantation [17]. Dialysis dependence is captured by Z99.2, and kidney transplant status is captured by Z94.0 [59]. It should be kept in mind that even after transplantation, some patients may still have CKD because the transplanted organ may not return the patient back to baseline kidney function; therefore, the N18 code for the patient's CKD stage should be used alongside the Z94.0. ESRD or kidney transplant status indicates a very high risk of CVD, and, as such, are generally classified as "CKM stage 4B." The care of patients experiencing ESRD or who have had a transplant, therefore, warrants a multidisciplinary team consisting of nephrologists, cardiologists, and other specialists to ensure coordinated care between all members involved in the patient's treatment [60]. The accurate coding for patients suffering from chronic renal failure is vital to ensure that specific specialized services are provided, and to track the outcomes of this population, which is classified as high risk.

Obesity and Overweight

Insulin resistance, T2DM, dyslipidemia, and hypertension are strongly influenced by obesity—particularly abdominal adiposity—which is a major upstream driver of CKM progression [61,62]. To enable stage-based identification and monitoring in real-world data, *ICD-10-CM* provides diagnosis codes for overweight and obesity and companion BMI Z-codes. Importantly, adult obesity severity can be coded using the obesity class codes E66.811 (class 1), E66.812 (class 2), and E66.813 (class 3) [17]. When available, companion BMI Z-codes may be used as optional supporting documentation to improve phenotypic granularity, but they are not required when an overweight or obesity diagnosis code is already present. These codes improve phenotyping granularity compared with legacy "morbid obesity" labels and support more consistent CKM stage 1 identification.

XSL•FO
RenderX

Adult Obesity Class 1-3 (E66.811-E66.813)

Adult obesity class codes map to BMI-based severity strata (class 1: BMI 30.0 - 34.9, class 2: BMI 35.0 - 39.9, and class 3: BMI≥40.0). Companion BMI Z-codes (Z68.3x-Z68.4x) may be used as optional supporting codes when available, but they are not required for stage 1 classification if an obesity diagnosis code has already been documented [17]. While legacy obesity diagnoses (eg, E66.9 and other E66 subcodes) remain in use, the class codes provide a more clinically relevant, less stigmatizing way to represent severity and can improve risk stratification and treatment planning. Within the CKM framework, higher obesity class is associated with greater risk of progression to hypertension, T2DM, CKD, and CVD, underscoring the importance of accurate and specific obesity documentation for early CKM intervention [63,64].

Overweight and BMI Documentation (Z68)

ICD-10-CM codes Z68.25-Z68.29 are assigned when a person has a BMI classified as "overweight" (25.0 - 29.9 kg/m²) [17,65]. Overweight is recognized as an important risk factor for the development of CKM syndrome, although it does not meet the definition of a "disease." These codes provide the ability to consistently document the identification of overweight individuals in the EHR data and allow population-level tracking and research on the effects of obesity. When an individual meets the CKM criteria with only 1 risk factor, their condition is classified as CKM stage 1. Once another risk factor develops, such as hypertension, high blood sugar, or dyslipidemia, the severity of their CKM classification will be elevated to CKM stage 2. Therefore, BMI documentation can improve phenotypic granularity and support prevention-oriented stratification when available, but stage 1 classification does not require a BMI Z-code if an overweight or obesity diagnosis code is already documented.

Metabolic Liver Disease

Nonalcoholic fatty liver disease (NAFLD) and its more severe inflammatory form, nonalcoholic steatohepatitis (NASH), have become well-established as part of the CKM syndrome [66]. These conditions are often referred to as "the hepatic manifestation of metabolic syndrome" and are closely linked with obesity, insulin resistance, and T2DM [67-69]. In the strict AHA-aligned staging algorithm presented here, NAFLD or NASH are treated as supplementary (risk-enhancing) conditions rather than stage-defining criteria and can be used for stratification and sensitivity analyses.

NAFLD (K76.0)

The *ICD-10-CM* classification for "Fatty (change of) liver, not elsewhere classified" is K76.0 [17,70]. Although this classification does not specifically diagnose NAFLD, it serves to identify the existence of a fatty liver in the absence of a more defined diagnosis such as NASH. NAFLD identifies a serious level of metabolic dysfunction throughout the body and places a significant determinant of the likelihood of future events involving the heart and kidneys [71,72]. In the CKM framework, NAFLD may be diagnosed along with CKM stage 2 or higher, indicating the need for very aggressive lifestyle modification and continued follow-up that incorporates the liver, heart, and kidney in concert with one another, rather than independently.

NASH (K75.81)

K75.81 is the code for the most advanced and inflammatory type of liver damage due to fat, known as NASH [17]. NASH causes liver inflammation and injury and can lead to scarring (fibrosis), liver failure (cirrhosis), and other health issues [73]. The presence of NASH indicates that the body has a greater amount of metabolic dysfunction than is seen with other forms of fatty liver disease and has a greater chance of developing systemic problems related to that severity of metabolic dysfunction [71,74]. When a patient is diagnosed with NASH, the likelihood that the patient will be classified as a high-risk patient on the CKM spectrum (usually CKM stage 2 or 3) and require an extensive treatment plan that manages the metabolic problems causing the NASH, as well as monitors for possible liver-related complications.

Other Metabolism-Related Diseases and Risk-Enhancing Conditions: *ICD-10-CM* Coding

In addition to CKM syndrome's core disorders, many other health conditions significantly increase risk (ie, "risk-enhancers") for CKM syndrome. They are dyslipidemia, metabolic syndrome, hyperuricemia, obstructive sleep apnea (OSA), polycystic ovarian syndrome (PCOS), and CKD-related secondary hyperparathyroidism (SHPT) [67,68,75-77]. The co-occurrence of these conditions may inform CKM risk stratification, but not all of them are stage-defining. In this framework, only diagnoses explicitly mapped in the staging algorithm are used for CKM stage assignment, whereas the others function as complementary risk-enhancing conditions.

Dyslipidemia and Lipoprotein Metabolism Disorders (E78)

Dyslipidemia is a major contributor to ASCVD and a core risk domain within the CKM syndrome [69]. In *ICD-10-CM*, most dyslipidemias and disorders of lipoprotein metabolism are captured under E78 (disorders of lipoprotein metabolism and other lipidemias), including specific codes for pure hypercholesterolemia (E78.0), pure hyperglyceridemia or hypertriglyceridemia (E78.1), mixed hyperlipidemia (E78.2), other hyperlipidemia (E78.4, including E78.41 and E78.49), and unspecified hyperlipidemia (E78.5) [17]. In the AHA CKM staging framework, hypertriglyceridemia (fasting triglycerides ≥135 mg/dL) is considered a stage 2 metabolic risk factor. Accordingly, pure hypertriglyceridemia (E78.1) is retained as the primary stage 2 code, and mixed hyperlipidemia (E78.2) may also be retained as an expanded claims-based proxy to capture mixed triglyceride-elevated phenotypes in administrative data. Broader lipid abnormalities remain generally treated as risk-enhancing rather than stage-defining conditions. Although E78.5 (hyperlipidemia, unspecified) is a billable code, more specific E78 codes should be used whenever the lipid phenotype is known, to improve clinical characterization and support more targeted lipid-lowering strategies [75,78]. The presence of dyslipidemia in conjunction with other CKM risk factors substantially increases future cardiovascular risk and underscores the need for timely risk-based intervention.

XSL•FO
RenderX

Metabolic Syndrome (E88.810)

Metabolic syndrome represents the clustering of multiple risk factors including central obesity, hyperglycemia, hypertension, and dyslipidemia [76]. It is an important precursor state and phenotype of CKM. *ICD-10-CM* provides a dedicated code for metabolic syndrome, that is, E88.810 [17]. This code can be used when at least 3 of the classic diagnostic criteria are met, which include increased waist circumference, elevated triglycerides, reduced high-density lipoprotein cholesterol, elevated BP, and elevated fasting glucose [79]. In the AHA CKM staging framework, the presence of metabolic syndrome serves as a key marker for progression to CKM stage 2. It is also a crucial code for population screening and for the automated identification of individuals at CKM stage 2 in EHRs.

Hyperuricemia and Gout (E79.0 and M10)

Hyperuricemia is interconnected with hypertension, CKD, insulin resistance, and obesity, and is now recognized as an independent CKM risk factor [80-82]. Urate crystals impair endothelial function, incite inflammation and oxidative stress, and accelerate atherosclerosis [83]. The *ICD-10-CM* code E79.0 is used for asymptomatic hyperuricemia, which is useful for identifying metabolically vulnerable yet oligo-symptomatic individuals. The codes M10 are used for gout, with subcategories for acute, chronic, and tophaceous presentations [17]. Hyperuricemia often accompanies CKM stages 1‑2, whereas gout—especially when coexisting with CKD or CVD—usually implies a higher CKM stage and cardio-renal risk.

Sleep Apnea Disorders (G47.30, G47.31, G47.33, and G47.39)

OSA, characterized by repetitive upper-airway collapse during sleep, is strongly associated with hypertension, arrhythmia, HF, and insulin resistance, and is a prevalent CKM comorbidity [77]. *ICD-10-CM* code G47.33 denotes OSA (adult or pediatric). Related sleep apnea codes include G47.30 (sleep apnea, unspecified), G47.31 (primary central sleep apnea), and G47.39 (other sleep apnea) [17]. Recognition and coding of OSA facilitates early identification of high-risk individuals in CKM stages 1‑2. Continuous positive airway pressure therapy improves sleep quality and may lower BP and CVD risk, rendering OSA a modifiable risk-enhancing factor [84].

PCOS (E28.2)

PCOS is one of the most common endocrine disorders in reproductive-age women, characterized by hyperandrogenism, anovulation, and insulin resistance, and is strongly linked to obesity, T2DM, dyslipidemia, and increased long-term CVD risk [85-87]. It is viewed as an early female-predominant CKM phenotype. The *ICD-10-CM* code E28.2 is used for polycystic ovary syndrome. Young women with PCOS frequently occupy CKM stages 1‑2 even before developing overt hypertension or diabetes [17]. Early coding and intervention (weight control, insulin-sensitizing measures, and lipid management) can delay progression to higher-risk stages.

CKD-Related SHPT of Renal Origin (N25.81)

SHPT of renal origin is a hallmark complication of moderate-to-advanced CKD, driven by phosphate retention, impaired calcitriol synthesis, and hypocalcemia, leading to parathyroid hyperplasia and excess parathyroid hormone secretion [68]. It contributes to mineral and bone disorder, vascular calcification, LVH, and heightened CVD risk. In *ICD-10-CM*, CKD-related SHPD is specifically coded as N25.81 and is typically documented alongside N18 to indicate CKD stage (eg, N18.4+N25.81) [17]. Clinically, the emergence of N25.81 usually reflects CKM stages 3‑4 and indicates the need for integrated management of CKD-mineral bone disorder and CVD.

CVD Coding in CKM Syndrome

CVD is the central clinical outcome of CKM syndrome and a key determinant of progression from CKM stage 3 (subclinical cardiovascular injury or very high-risk status) to CKM stage 4 (overt clinical events) [88]. In the AHA framework, stage 3 commonly corresponds to subclinical atherosclerosis and/or early cardiac dysfunction, whereas stage 4 denotes established clinical ASCVD, HF, disabling stroke, and other “hard” end-stage events [1,2]. Accordingly, systematic mapping of major cardiovascular phenotypes and their *ICD-10* or *ICD-10-CM* codes is critical for identifying CKM stage 3‑4 populations in EHRs and epidemiological databases.

Coronary Heart Disease and Angina (I20 and I25)

Coronary heart disease encompasses a group of conditions caused by coronary atherosclerosis with luminal stenosis or occlusion, resulting in myocardial ischemia, hypoxia, and, in severe cases, necrosis [89-91]. Within the CKM paradigm, coronary disease is often the downstream “end point manifestation” after prolonged exposure to obesity, diabetes, hypertension, dyslipidemia, and CKD, and typically aligns with CKM stage 4 [57]. In *ICD-10-CM*, chronic atherosclerotic coronary disease is primarily coded under I25. Specifically, I25.10 denotes atherosclerotic heart disease of native coronary artery without angina pectoris, while native-vessel atherosclerotic disease with angina pectoris is captured under I25.11 (child codes I25.110-I25.119) [17]. Atherosclerosis of coronary artery bypass grafts or of a coronary artery of a transplanted heart with angina pectoris is captured under I25.7, whereas bypass-graft atherosclerosis of coronary artery bypass grafts without angina pectoris can be captured under I25.810. When angina is clinically documented, angina pectoris is coded under I20 (eg, I20.0 unstable angina, I20.1 angina with documented spasm, I20.2 refractory angina, I20.8 other forms, and I20.9 unspecified). Co-occurrence of I25 with E11 (diabetes), I10-I13 (hypertension), and N18 (CKD) often indicates advanced, multiorgan CKM with very high event and mortality risk, and is frequently used to define major adverse cardiovascular events in research settings.

AMI (I21 and I22)

AMI is the most severe acute presentation of coronary disease, typically triggered by rupture of a vulnerable plaque with superimposed thrombosis, leading to complete or near-complete coronary occlusion and long-term myocardial necrosis [92]. In CKM populations, obesity, diabetes, and CKD increase plaque burden and vulnerability, markedly raising AMI risk [93]. *ICD-10-CM* classifies AMI under I21, including I21.0-I21.3

XSL•FO
RenderX

for ST-segment elevation myocardial infarction by site or wall and I21.4 for non–ST-segment elevation myocardial infarction; I21.A1 specifies myocardial infarction type 2 (supply-demand mismatch), which may occur in contexts such as severe anemia, tachyarrhythmia, or hypertensive crisis. I22 captures subsequent myocardial infarction (MI) occurring within 4 weeks (28 d) of a previous acute MI, regardless of site [17]. In CKM staging and clinical research, I21 and I22 are typically treated as "hard end points"; the presence of these codes generally indicates CKM stage 4 and warrants intensive long-term secondary prevention (eg, high-intensity lipid lowering, antiplatelet therapy, and strict BP and glycemic control) [94].

HF (I50, I11.0, and I13)

HF represents the common final pathway of many structural and functional cardiac disorders, characterized by impaired ventricular systolic and/or diastolic function such that cardiac output fails to meet metabolic demands (or is maintained only at the expense of elevated filling pressures) [94,95]. In CKM syndrome, hypertensive remodeling, diabetic cardiomyopathy, ischemic cardiomyopathy, and CKD-related volume overload all contribute to HF development. *ICD-10-CM* codes for HF include I50.1 (left ventricular failure), I50.2 (systolic [congestive] HF), I50.3 (diastolic [congestive] HF), I50.4 (combined systolic and diastolic HF), and I50.9 (HF, unspecified). When HF is explicitly attributed to long-standing hypertension, I11.0 (hypertensive heart disease with HF) should be used, with an additional code from I50. to identify the type of HF. When hypertensive heart disease coexists with CKD and HF, I13.0 (with CKD stage 1 - 4 or unspecified CKD) or I13.2 (with CKD stage 5 or ESRD) is used, with additional codes from I50 (HF type) and N18 (CKD stage). In CKM staging, the presence of I50 (or I11.0 with I50; or I13.0 or I13.2 with I50 and N18) generally places patients in CKM stage 4 (clinical HF), representing the highest cardio-renal comorbidity burden [1,2,17].

Arrhythmias, With a Focus on Atrial Fibrillation (I48)

Atrial fibrillation (AF) is the most common long-term arrhythmia, characterized by rapid, disorganized atrial electrical activity, loss of effective atrial contraction, blood stasis, and a propensity for mural thrombus formation—making it a major cause of ischemic stroke [96]. Obesity, OSA, hypertension, T2DM, and CKD promote atrial structural and electrical remodeling, increasing AF risk [77,96]. In *ICD-10-CM*, AF and flutter are coded under I48, including I48.0 (paroxysmal AF), I48.19 (persistent AF), I48.20 (chronic AF), and I48.21 (permanent AF) [17]. Other arrhythmias common in advanced CKM include ventricular tachycardia (I47.2), premature ventricular contractions (I49.3), and ventricular fibrillation (I49.0). AF can be regarded as a CVD-equivalent condition; when combined with previous stroke (I63) or HF (I50), it typically indicates the highest-risk tier and aligns with CKM stage 4.

PAD (I70 and I73)

PAD refers to atherosclerotic narrowing or occlusion of medium-to-large arteries outside the coronary and intracranial circulations, most commonly affecting the lower extremities [77]. It is a marker of systemic atherosclerosis and often coexists with coronary and cerebrovascular disease [97]. *ICD-10-CM* codes include I70.2 (atherosclerosis of native arteries of the extremities) and I73.9 (peripheral vascular disease, unspecified), among others [17]. In CKM staging, PAD is generally categorized as clinical ASCVD; the presence of I70 or I73 typically indicates CKM stage 4.

Stroke (I60-I64)

Stroke, including ischemic and hemorrhagic subtypes, is among the most severe and disabling end points across the CKM spectrum. CKM individuals often exhibit clustering of hypertension, diabetes, AF, and dyslipidemia, resulting in cumulative cerebrovascular risk. In *ICD-10-CM*, acute stroke events are generally captured under I60-I64, including I60 (nontraumatic subarachnoid hemorrhage), I61 (nontraumatic intracerebral hemorrhage), I62 (other and unspecified nontraumatic intracranial hemorrhage), I63 (cerebral infarction or ischemic stroke), and I64 (stroke, not specified as hemorrhage or infarction) [17]. To improve ascertainment of previous clinical cerebrovascular disease in claims or EHR data, history or sequelae codes can also be captured, including I69 (sequelae of cerebrovascular disease) and Z86.73 (personal history of transient ischemic attack and cerebral infarction without residual deficits) [17]. The occurrence of stroke indicates CKM stage 4 and is frequently associated with substantial functional impairment and reduced quality of life, serving as a key end point for evaluating integrated cardio-renal-metabolic prevention.

Discussion

Principal Findings

In this study, we developed an AHA-aligned *ICD-10-CM* coding framework to operationalize CKM syndrome stages 0 - 4 in EHRs and claims data. The framework provides stage-specific code sets, a hierarchical staging algorithm, and practical rules for co-occurrence, lookback periods, and encounter confirmation. These components provide a transparent and reproducible approach for CKM staging in real-world datasets.

Clinical and Research Implications of a Standardized Coding Framework

Establishing a consistent *ICD-10-CM* coding framework for CKM has important clinical and research implications. Consistent coding can improve the completeness, transparency, and interpretability of patient-related data for clinicians, while supporting cross-specialty communication and the development of stage-aware decision support tools [98]. These tools will allow clinicians to automatically identify the CKM stage for their patients based on a coded diagnosis, which will trigger a clinician's response to provide stage-specific interventions [99]. For instance, clinicians could provide lifestyle counseling to patients diagnosed with CKM stage 1 and intensive management of risk factors to patients diagnosed with CKM stage 3. On the research side, a standardized coding framework is necessary for conducting large, multicenter studies on the epidemiology, natural history, and treatment of CKM syndrome [89]. The use of a standardized code system can support more reproducible

XSL•FO
RenderX

population identification, more transparent exposure and outcome ascertainment, and more consistent comparison across studies and health care systems [40,100].

Framework Validity Scope and Comparison With Existing *ICD*-Based CKM Staging Approaches

The validity of any *ICD*-based staging framework rests on (1) content validity—whether included codes represent the stage-defining phenotypes, (2) structural validity—whether the algorithm assigns mutually exclusive stages while preserving clinical hierarchy, and (3) operational validity—whether implementation choices (eg, ascertainment windows and encounter-confirmation rules) improve reproducibility, transparency, and consistency of stage assignment across datasets.

We enhanced content validity by anchoring every stage to the AHA definitions and prioritizing granular *ICD-10-CM* codes and etiologic combination codes for kidney and cardiovascular conditions. All codes and ranges were standardized against current US *ICD-10-CM* conventions (FY2026) to avoid non-*ICD* labels and ambiguous shorthand, and we explicitly note newer codes that may be absent from older datasets (eg, E11.A; adult obesity class codes E66.811-E66.813). Structural validity is addressed through an explicit top-down algorithm that assigns the highest qualifying stage (4->0), which prevents double counting and makes precedence rules transparent.

Previously published *ICD*-based operationalizations vary and often omit key design decisions that affect classification consistency and implementation reproducibility [101]. Common limitations include incomplete capture of adiposity-only stage 1, inconsistent handling of subclinical stage 3 (often merged with stage 2 in claims data), and insufficiently specified precedence, timing, and encounter-confirmation rules [93]. In contrast, our framework (1) provides modular, stage-specific code domains; (2) separates claims-only staging from EHR-enhanced staging; and (3) operationalizes stage 3 explicitly by distinguishing an EHR-defined subclinical CVD component from a conservative claims-only definition based on very-high risk CKD. These choices improve transparency and facilitate sensitivity analyses when the available data cannot fully represent subclinical disease.

With respect to validation, the present work provides literature- and guideline-based validation (code verification and crosswalk to published CKM algorithms) rather than empirical performance metrics. We therefore recommend that future external evaluations benchmark the framework in specific target datasets using chart review, registry linkage, or structured EHR reference standards under prespecified encounter-confirmation and lookback rules [40,90,91].

Facilitating Multidisciplinary and Coordinated Care

The CKM framework includes a focus on an integrated team-based approach to care instead of fragmented siloed care, with standardized coding using the *ICD-10-CM* for coding conditions related to CKD [102,103]. Using a shared code set across providers creates a common representation of the patient's condition, health status, and risk profile. This supports cross-specialty communication and coordination around an integrated metabolic-renal-cardiovascular care plan. A shared code set standardizes the patient's risk profile and supports cross-specialty coordination. For example, documenting diabetic nephropathy (E11.22) and stage 3 CKD (N18.3) can prompt proactive cardiology communication given the elevated cardiovascular risk. This coordination is central to improving outcomes in advanced CKM.

Limitations and Challenges of *ICD-10-CM* Coding for CKM

Although *ICD-10-CM* codes enable scalable CKM case identification, several limitations should be considered. First, coding accuracy depends on the completeness and specificity of clinical documentation; underdocumentation may lead to underascertainment of CKM phenotypes and misclassification of stage [104]. Second, *ICD-10-CM* may not fully capture etiologic nuance (eg, differentiating diabetic vs hypertensive CKD when documentation is ambiguous), which can introduce heterogeneity in stage definitions across sites [105]. Third, stage 3 is inherently challenging to operationalize in claims-only data because subclinical disease is often defined by imaging findings (eg, coronary artery calcium) or biomarkers (eg, high-sensitivity troponin or natriuretic peptides) that are not reliably represented in diagnosis codes. Accordingly, stage 3 identification may require an EHR-enhanced approach (integrating structured laboratories or imaging results) or should be interpreted as a conservative proxy when using claims-only algorithms [106,107]. Finally, because *ICD-10-CM* is updated annually, code sets should be periodically reviewed and maintained to remain consistent with current conventions and guidance [108].

Future Directions and Recommendations

Going forward, there are several basic actions that need to be taken to further develop and use a standardized *ICD-10-CM* coding framework for CKM syndrome. First, validation studies should be conducted to evaluate the coding algorithm's performance when used in different types of health care settings and patient populations [109]. Validation studies will compare the CKM stages assigned through this automated coding framework to CKM stages assigned using traditional medical record review methods, allowing for a comparison of the accuracy and dependability between automated methods and traditional methods of determining to assign CKM stages [110]. Second, CKM staging frameworks should be incorporated into established clinical guidelines and evaluated in quality metrics. An increased awareness of the CKM syndrome among health care providers and the increased use of a standardized coding system should be an outcome of this effort. Additionally, ongoing education and training for health care providers, medical coders, and others associated with health care should be provided regarding the importance of accurate, specific, and complete CKM documentation and codification of CKM-related conditions [111]. Education and training could be provided in a variety of formats, including online training modules, workshops, and various types of educational materials. Last but not least, as CKM research progresses, the *ICD-10-CM* coding framework should be periodically reviewed for consistency with current CKM findings and be updated to include new codes as they are available [112]. The achievement of these objectives

XSL•FO
RenderX

is likely to result in continued relevance and improvement of the *ICD-10-CM* coding framework for CKM syndrome. In addition, as natural language processing and information extraction mature, subclinical findings documented in unstructured EHR notes (eg, coronary artery calcium reporting, echocardiography interpretations, or biomarker results referenced in narratives) could be leveraged to improve stage 3 ascertainment when structured data are incomplete. Finally, integrating this staging framework into clinical quality measures and electronic decision support tools may facilitate scalable, stage-appropriate prevention and care pathways while enabling ongoing monitoring and iterative refinement of the code sets over time.

Conclusion

The CKM syndrome is a growing public health concern because metabolic dysfunction, CKD, and CVD interact bidirectionally and can amplify one another over time. The AHA proposed a staged CKM framework (0 - 4) to support risk stratification and stage-appropriate intervention, but the absence of a single *ICD-10-CM* code has limited standardized implementation in real-world data. In this manuscript, we provide an AHA-aligned *ICD-10-CM* operational framework that links stage-defining phenotypes and outcomes to code sets and reproducible algorithmic rules. If adopted and maintained, this framework can support consistent cohort identification, risk stratification, and outcome assessment across health systems and claims databases while enabling future empirical validation and iterative refinement.

Funding

This work was supported by the National Key R&D Program of China (2021YFC2500501) and the National Natural Science Foundation of China (U25D9022).

Data Availability

No individual-level patient data were generated or analyzed for this framework-development study; therefore, no underlying participant-level dataset is available for sharing.

Authors' Contributions

Methodology, investigation, formal analysis, writing – original draft, writing – review and editing: MZ, JZ.
Investigation, formal analysis: HW, CX, JL, CL.
Conceptualization, funding acquisition, supervision: XH, XZ.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information.
[ZIP File, 12 KB - diabetes_v11i1e91827_app1.zip]

References

1. Ndumele CE, Rangaswami J, Chow SL, et al. Cardiovascular-kidney-metabolic health: a presidential advisory from the American Heart Association. Circulation 2023 Nov 14;148(20):1606-1635. [doi: 10.1161/CIR.0000000000001184] [Medline: 37807924]
2. Ndumele CE, Neeland IJ, Tuttle KR, et al. A synopsis of the evidence for the science and clinical management of cardiovascular-kidney-metabolic (CKM) syndrome: a scientific statement from the American Heart Association. Circulation 2023 Nov 14;148(20):1636-1664. [doi: 10.1161/CIR.0000000000001186] [Medline: 37807920]
3. Zuin M, Aggarwal R, Khera A, Hermes Z, Bilato C, Piazza G. Cardiovascular-kidney-metabolic syndrome-attributable mortality in the United States, from 2010 to 2023. JACC Adv 2025 Dec;4(12 Pt 2):102369. [doi: 10.1016/j.jacadv.2025.102369] [Medline: 41270306]
4. Ostrominski JW, Arnold SV, Butler J, et al. Prevalence and overlap of cardiac, renal, and metabolic conditions in US adults, 1999-2020. JAMA Cardiol 2023 Nov 1;8(11):1050-1060. [doi: 10.1001/jamacardio.2023.3241] [Medline: 37755728]
5. Aggarwal R, Ostrominski JW, Vaduganathan M. Prevalence of cardiovascular-kidney-metabolic syndrome stages in US adults, 2011-2020. JAMA 2024 Jun 4;331(21):1858-1860. [doi: 10.1001/jama.2024.6892] [Medline: 38717747]
6. Iacoviello M, Gori M, Grandaliano G, Minutolo R, Pitocco D, Trevisan R. A holistic approach to managing cardio-kidney metabolic syndrome: insights and recommendations from the Italian perspective. Front Cardiovasc Med 2025;12:1583702. [doi: 10.3389/fcvm.2025.1583702] [Medline: 40264513]
7. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61-81. [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]

XSL•FO
RenderX

8. Abul-Husn NS, Kenny EE. Personalized medicine and the power of electronic health records. Cell 2019 Mar 21;177(1):58-69. [doi: 10.1016/j.cell.2019.02.039] [Medline: 30901549]
9. Shen Y, Yu J, Zhou J, Hu G. Twenty-five years of evolution and hurdles in electronic health records and interoperability in medical research: comprehensive review. J Med Internet Res 2025 Jan 9;27:e59024. [doi: 10.2196/59024] [Medline: 39787599]
10. Marsolo KA, Cheville A, Melnick ER, et al. Impact of electronic health record updates and changes on the delivery and monitoring of interventions in embedded pragmatic clinical trials. Contemp Clin Trials 2025 Jan;148:107744. [doi: 10.1016/j.cct.2024.107744] [Medline: 39561918]
11. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol 2021 Oct 27;21(1):234. [doi: 10.1186/s12874-021-01416-5] [Medline: 34706667]
12. Matthewman J, Andresen K, Suffel A, et al. Checklist and guidance on creating codelists for routinely collected health data research. NIHR Open Res 2024;4:20. [doi: 10.3310/nihropenres.13550.2] [Medline: 39345273]
13. Alonso-Morán E, Nuño-Solinis R, Onder G, Tonnara G. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. Eur J Intern Med 2015 Apr;26(3):182-189. [doi: 10.1016/j.ejim.2015.02.010] [Medline: 25753935]
14. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc 2016 Nov;23(6):1046-1052. [doi: 10.1093/jamia/ocv202] [Medline: 27026615]
15. Orozco-Beltrán D, Quiroga B, Esteban-Fernández A, et al. Evaluation, management and therapeutic approach of cardiovascular-kidney-metabolic syndrome: a multidisciplinary delphi expert consensus. J Clin Med 2025 Dec 17;14(24):8930. [doi: 10.3390/jcm14248930] [Medline: 41464833]
16. Pacheco JA, Rasmussen LV, Wiley K Jr, et al. Evaluation of the portability of computable phenotypes with natural language processing in the eMERGE network. Sci Rep 2023 Feb 3;13(1):1971. [doi: 10.1038/s41598-023-27481-y] [Medline: 36737471]
17. ICD-10-CM Official Guidelines for Coding and Reporting: FY 2026 -- Updated October 1, 2025: (October 1, 2025 - September 30, 2026). : U.S. Centers for Disease Control and Prevention; 2025 URL: https://stacks.cdc.gov/view/cdc/250974 [accessed 2026-05-15]
18. Writing Committee Members*, Jones DW, Ferdinand KC, et al. 2025 AHA/ACC/AANP/AAPA/ABC/ACCP/ACPM/AGS/AMA/ASPC/NMA/PCNA/SGIM guideline for the prevention, detection, evaluation and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2025 Sep 16;152(11):e114-e218. [doi: 10.1161/CIR.0000000000001356] [Medline: 40811497]
19. Bajaj M, McCoy RG, Balapattabi K, et al. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of care in diabetes—2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S61-S88. [doi: 10.2337/dc26-S004] [Medline: 41854433]
20. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int 2024 Apr;105(4S):S117-S314. [doi: 10.1016/j.kint.2023.10.018] [Medline: 38490803]
21. Elmaleh-Sachs A, Schwartz JL, Bramante CT, Nicklas JM, Gudzune KA, Jay M. Obesity management in adults: a review. JAMA 2023 Nov 28;330(20):2000-2015. [doi: 10.1001/jama.2023.19897] [Medline: 38015216]
22. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO clinical practice guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol 2024 Sep;81(3):492-542. [doi: 10.1016/j.jhep.2024.04.031] [Medline: 38851997]
23. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980. [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
24. Fuchs FD, Whelton PK. High blood pressure and cardiovascular disease. Hypertension 2020 Feb;75(2):285-292. [doi: 10.1161/HYPERTENSIONAHA.119.14240] [Medline: 31865786]
25. Đorđević DB, Koračević GP, Đorđević AD, Lović DB. Hypertension and left ventricular hypertrophy. J Hypertens 2024 Sep 1;42(9):1505-1515. [doi: 10.1097/HJH.0000000000003774] [Medline: 38747417]
26. Messerli FH, Rimoldi SF, Bangalore S. The transition from hypertension to heart failure: contemporary update. JACC Heart Fail 2017 Aug;5(8):543-551. [doi: 10.1016/j.jchf.2017.04.012] [Medline: 28711447]
27. Luo D, Cheng Y, Zhang H, et al. Association between high blood pressure and long term cardiovascular events in young adults: systematic review and meta-analysis. BMJ 2020 Sep 9;370:m3222. [doi: 10.1136/bmj.m3222] [Medline: 32907799]
28. Poznyak AV, Sadykhov NK, Kartuesov AG, et al. Hypertension as a risk factor for atherosclerosis: cardiovascular risk assessment. Front Cardiovasc Med 2022;9:959285. [doi: 10.3389/fcvm.2022.959285] [Medline: 36072873]
29. Burnier M, Damianaki A. Hypertension as cardiovascular risk factor in chronic kidney disease. Circ Res 2023 Apr 14;132(8):1050-1063. [doi: 10.1161/CIRCRESAHA.122.321762] [Medline: 37053276]

XSL•FO
RenderX

30. Rahimi K, Bidel Z, Nazarzadeh M, et al. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. The Lancet 2021 May;397(10285):1625-1636. [doi: 10.1016/S0140-6736(21)00590-0]
31. Muntner P, Shimbo D, Carey RM, et al. Measurement of blood pressure in humans: a scientific statement from the American Heart Association. Hypertension 2019 May;73(5):e35-e66. [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]
32. He S, Park S, Kuklina E, et al. Leveraging electronic health records to construct a phenotype for hypertension surveillance in the United States. Am J Hypertens 2023 Nov 15;36(12):677-685. [doi: 10.1093/ajh/hpad081] [Medline: 37696605]
33. Song J, Wang X, Wang B, et al. Learning implementation of a guideline based decision support system to improve hypertension treatment in primary care in China: pragmatic cluster randomised controlled trial. BMJ 2024 Jul 23;386:e079143. [doi: 10.1136/bmj-2023-079143] [Medline: 39043397]
34. Phadke A, Weng Y, Johnson CB, et al. Integrating a high blood pressure advisory across a primary care network. JAMA Netw Open 2025 Apr 1;8(4):e257313. [doi: 10.1001/jamanetworkopen.2025.7313] [Medline: 40279123]
35. Masenga SK, Kirabo A. Hypertensive heart disease: risk factors, complications and mechanisms. Front Cardiovasc Med 2023;10:1205475. [doi: 10.3389/fcvm.2023.1205475] [Medline: 37342440]
36. Díez J, Butler J. Growing heart failure burden of hypertensive heart disease: a call to action. Hypertension 2023 Jan;80(1):13-21. [doi: 10.1161/HYPERTENSIONAHA.122.19373] [Medline: 36082670]
37. Inamdar AA, Inamdar AC. Heart failure: diagnosis, management and utilization. J Clin Med 2016 Jun 29;5(7):62. [doi: 10.3390/jcm5070062] [Medline: 27367736]
38. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022 May 3;145(18):e895-e1032. [doi: 10.1161/CIR.0000000000001063] [Medline: 35363499]
39. Marwick TH, Gillebert TC, Aurigemma G, et al. Recommendations on the use of echocardiography in adult hypertension: a report from the European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE). J Am Soc Echocardiogr 2015 Jul;28(7):727-754. [doi: 10.1016/j.echo.2015.05.002] [Medline: 26140936]
40. Bates BA, Akhabue E, Nahass MM, et al. Validity of International Classification of Diseases (ICD)-10 diagnosis codes for identification of acute heart failure hospitalization and heart failure with reduced versus preserved ejection fraction in a National Medicare sample. Circ Cardiovasc Qual Outcomes 2023 Feb;16(2):e009078. [doi: 10.1161/CIRCOUTCOMES.122.009078] [Medline: 36688301]
41. Samal L, Kilgallon JL, Lipsitz S, et al. Clinical decision support for hypertension management in chronic kidney disease: a randomized clinical trial. JAMA Intern Med 2024 May 1;184(5):484-492. [doi: 10.1001/jamainternmed.2023.8315] [Medline: 38466302]
42. Ku E, Lee BJ, Wei J, Weir MR. Hypertension in CKD: core curriculum 2019. Am J Kidney Dis 2019 Jul;74(1):120-131. [doi: 10.1053/j.ajkd.2018.12.044] [Medline: 30898362]
43. Hallan SI, Øvrehus MA, Bjørneklett R, Aasarød KI, Fogo AB, Ix JH. Hypertensive nephrosclerosis: wider kidney biopsy indications may be needed to improve diagnostics. J Intern Med 2021 Jan;289(1):69-83. [doi: 10.1111/joim.13146] [Medline: 32613703]
44. Francis A, Harhay MN, Ong ACM, et al. Chronic kidney disease and the global public health agenda: an international consensus. Nat Rev Nephrol 2024 Jul;20(7):473-485. [doi: 10.1038/s41581-024-00820-6] [Medline: 38570631]
45. Rangaswami J, Bhalla V, Blair JEA, et al. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the American Heart Association. Circulation 2019 Apr 16;139(16):e840-e878. [doi: 10.1161/CIR.0000000000000664] [Medline: 30852913]
46. Méndez AB, Azancot MA, Olivella A, Soler MJ. New aspects in cardiorenal syndrome and HFpEF. Clin Kidney J 2022 Oct;15(10):1807-1815. [doi: 10.1093/ckj/sfac133] [Medline: 36158149]
47. Ryan DK, Banerjee D, Jouhra F. Management of heart failure in patients with chronic kidney disease. Eur Cardiol 2022 Feb;17:e17. [doi: 10.15420/ecr.2021.33] [Medline: 35990402]
48. Linde C, Bakhai A, Furuland H, et al. Real-world associations of renin-angiotensin-aldosterone system inhibitor dose, hyperkalemia, and adverse clinical outcomes in a cohort of patients with new-onset chronic kidney disease or heart failure in the United Kingdom. J Am Heart Assoc 2019 Nov 19;8(22):e012655. [doi: 10.1161/JAHA.119.012655] [Medline: 31711387]
49. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet 2014 Mar 22;383(9922):1068-1083. [doi: 10.1016/S0140-6736(13)62154-6] [Medline: 24315620]
50. Martinez Leon V, Hilburg R, Susztak K. Mechanisms of diabetic kidney disease and established and emerging treatments. Nat Rev Endocrinol 2026 Jan;22(1):21-35. [doi: 10.1038/s41574-025-01171-3] [Medline: 40935879]
51. Rawshani A, Rawshani A, Franzén S, et al. Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2018 Aug 16;379(7):633-644. [doi: 10.1056/NEJMoa1800256] [Medline: 30110583]
52. Dugan J, Shubrook J. International Classification of Diseases, 10th Revision, Coding for Diabetes. Clin Diabetes 2017 Oct;35(4):232-238. [doi: 10.2337/cd16-0052] [Medline: 29109613]
53. Bajaj M, McCoy RG, Balapattabi K, et al. 2. Diagnosis and classification of diabetes: standards of care in diabetes—2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S27-S49. [doi: 10.2337/dc26-S002] [Medline: 41854433]

XSL•FO
RenderX

54. Bajaj M, McCoy RG, Balapattabi K, et al. 3. Prevention or delay of diabetes and associated comorbidities: standards of care in diabetes—2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S50-S60. [doi: 10.2337/dc26-S003] [Medline: 41854433]
55. Riddle MC, Cefalu WT, Evans PH, et al. Consensus report: definition and interpretation of remission in type 2 diabetes. Diabetes Care 2021 Aug 30;44(10):2438-2444. [doi: 10.2337/dci21-0034] [Medline: 34462270]
56. American Diabetes Association Professional Practice Committee for Diabetes*. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes—2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S89-S131. [doi: 10.2337/dc26-S005] [Medline: 41358898]
57. Jankowski J, Floege J, Fliser D, Böhm M, Marx N. Cardiovascular disease in chronic kidney disease: pathophysiological Insights and therapeutic options. Circulation 2021 Mar 16;143(11):1157-1172. [doi: 10.1161/CIRCULATIONAHA.120.050686] [Medline: 33720773]
58. Chase HS, Radhakrishnan J, Shirazian S, Rao MK, Vawdrey DK. Under-documentation of chronic kidney disease in the electronic health record in outpatients. J Am Med Inform Assoc 2010;17(5):588-594. [doi: 10.1136/jamia.2009.001396] [Medline: 20819869]
59. Chapter 1: definition and classification of CKD. Kidney Int Suppl (2011) 2013 Jan;3(1):19-62. [doi: 10.1038/kisup.2012.64] [Medline: 25018975]
60. Birdwell KA, Park M. Post-transplant cardiovascular disease. Clin J Am Soc Nephrol 2021 Dec;16(12):1878-1889. [doi: 10.2215/CJN.00520121] [Medline: 34556500]
61. Heymsfield SB, Wadden TA. Mechanisms, pathophysiology, and management of obesity. N Engl J Med 2017 Jan 19;376(3):254-266. [doi: 10.1056/NEJMra1514009] [Medline: 28099824]
62. Kessler C. Pathophysiology of obesity. Nurs Clin North Am 2021 Dec;56(4):465-478. [doi: 10.1016/j.cnur.2021.08.001] [Medline: 34749888]
63. Jayedi A, Soltani S, Motlagh SZT, et al. Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. BMJ 2022 Jan 18;376:e067516. [doi: 10.1136/bmj-2021-067516] [Medline: 35042741]
64. Jayedi A, Rashidy-Pour A, Khorshidi M, Shab-Bidar S. Body mass index, abdominal adiposity, weight gain and risk of developing hypertension: a systematic review and dose-response meta-analysis of more than 2.3 million participants. Obes Rev 2018 May;19(5):654-667. [doi: 10.1111/obr.12656] [Medline: 29334692]
65. Suissa K, Schneeweiss S, Lin KJ, Brill G, Kim SC, Patorno E. Validation of obesity-related diagnosis codes in claims data. Diabetes Obes Metab 2021 Dec;23(12):2623-2631. [doi: 10.1111/dom.14512] [Medline: 34338404]
66. Duell PB, Welty FK, Miller M, et al. Nonalcoholic fatty liver disease and cardiovascular risk: a scientific statement from the American Heart Association. Arterioscler Thromb Vasc Biol 2022 Jun;42(6):e168-e185. [doi: 10.1161/ATV.0000000000000153] [Medline: 35418240]
67. Teede HJ, Misso ML, Costello MF, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Fertil Steril 2018 Aug;110(3):364-379. [doi: 10.1016/j.fertnstert.2018.05.004] [Medline: 30033227]
68. KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease–Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl (2011) 2017 Jul;7(1):1-59. [doi: 10.1016/j.kisu.2017.04.001]
69. Borén J, Chapman MJ, Krauss RM, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease: pathophysiological, genetic, and therapeutic insights: a consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J 2020 Jun 21;41(24):2313-2330. [doi: 10.1093/eurheartj/ehz962] [Medline: 32052833]
70. Hagström H, Adams LA, Allen AM, et al. Administrative coding in electronic health care record-based research of NAFLD: an expert panel consensus statement. Hepatology 2021 Jul;74(1):474-482. [doi: 10.1002/hep.31726] [Medline: 33486773]
71. Mantovani A, Csermely A, Petracca G, et al. Non-alcoholic fatty liver disease and risk of fatal and non-fatal cardiovascular events: an updated systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2021 Nov;6(11):903-913. [doi: 10.1016/S2468-1253(21)00308-3] [Medline: 34555346]
72. Mantovani A, Petracca G, Beatrice G, et al. Non-alcoholic fatty liver disease and risk of incident chronic kidney disease: an updated meta-analysis. Gut 2022 Jan;71(1):156-162. [doi: 10.1136/gutjnl-2020-323082] [Medline: 33303564]
73. Powell EE, Wong VWS, Rinella M. Non-alcoholic fatty liver disease. Lancet 2021 Jun 5;397(10290):2212-2224. [doi: 10.1016/S0140-6736(20)32511-3] [Medline: 33894145]
74. Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med 2021 Oct 21;385(17):1559-1569. [doi: 10.1056/NEJMoa2029349] [Medline: 34670043]
75. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Eur Heart J 2020 Jan 1;41(1):111-188. [doi: 10.1093/eurheartj/ehz455] [Medline: 31504418]
76. Alberti K, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the

XSL•FO
RenderX

Study of Obesity. Circulation 2009 Oct 20;120(16):1640-1645. [doi: 10.1161/CIRCULATIONAHA.109.192644] [Medline: 19805654]
77. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive sleep apnea and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2021 Jul 20;144(3):e56-e67. [doi: 10.1161/CIR.0000000000000988] [Medline: 34148375]
78. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Jun 18;139(25):e1082-e1143. [doi: 10.1161/CIR.0000000000000625] [Medline: 30586774]
79. Grundy SM, Cleeman JI, Daniels SR, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005 Oct 25;112(17):2735-2752. [doi: 10.1161/CIRCULATIONAHA.105.169404] [Medline: 16157765]
80. Borghi C, Agnoletti D, Cicero AFG, Lurbe E, Virdis A. Uric acid and hypertension: a review of evidence and future perspectives for the management of cardiovascular risk. Hypertension 2022 Sep;79(9):1927-1936. [doi: 10.1161/HYPERTENSIONAHA.122.17956] [Medline: 35658505]
81. Gonçalves DLN, Moreira TR, da Silva LS. A systematic review and meta-analysis of the association between uric acid levels and chronic kidney disease. Sci Rep 2022 Apr 15;12(1):6251. [doi: 10.1038/s41598-022-10118-x] [Medline: 35428828]
82. Copur S, Demiray A, Kanbay M. Uric acid in metabolic syndrome: does uric acid have a definitive role? Eur J Intern Med 2022 Sep;103:4-12. [doi: 10.1016/j.ejim.2022.04.022] [Medline: 35508444]
83. Kimura Y, Tsukui D, Kono H. Uric acid in inflammation and the pathogenesis of atherosclerosis. Int J Mol Sci 2021 Nov 17;22(22):12394. [doi: 10.3390/ijms222212394] [Medline: 34830282]
84. Yu J, Zhou Z, McEvoy RD, et al. Association of positive airway pressure with cardiovascular events and death in adults with sleep apnea: a systematic review and meta-analysis. JAMA 2017 Jul 11;318(2):156-166. [doi: 10.1001/jama.2017.7967] [Medline: 28697252]
85. Tay CT, Mousa A, Vyas A, Pattuwage L, Tehrani FR, Teede H. 2023 International Evidence-Based Polycystic Ovary Syndrome Guideline update: insights from a systematic review and meta-analysis on elevated clinical cardiovascular disease in polycystic ovary syndrome. J Am Heart Assoc 2024 Aug 20;13(16):e033572. [doi: 10.1161/JAHA.123.033572] [Medline: 39119982]
86. Lim SS, Kakoly NS, Tan JWJ, et al. Metabolic syndrome in polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. Obes Rev 2019 Feb;20(2):339-352. [doi: 10.1111/obr.12762] [Medline: 30339316]
87. Wekker V, van Dammen L, Koning A, et al. Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. Hum Reprod Update 2020 Nov 1;26(6):942-960. [doi: 10.1093/humupd/dmaa029] [Medline: 32995872]
88. Cheng Z, Xu W, Luo J, et al. Using the American Heart Association's new "cardiovascular-kidney-metabolic syndrome" metrics to assess life expectancy in US adults: a life table analysis. Am J Prev Cardiol 2025 Dec;24:101339. [doi: 10.1016/j.ajpc.2025.101339] [Medline: 41256040]
89. Williams R, Kontopantelis E, Buchan I, Peek N. Clinical code set engineering for reusing EHR data for research: a review. J Biomed Inform 2017 Jun;70(1-13):1-13. [doi: 10.1016/j.jbi.2017.04.010] [Medline: 28442434]
90. Taya S, Ejiri K, Kaneko H, et al. Sex differences in the progression of cardiovascular-kidney-metabolic syndrome. Eur Heart J Open 2025 Nov;5(6):oeaf162. [doi: 10.1093/ehjopen/oeaf162] [Medline: 41451240]
91. Kim BS, Kim HJ, Kim H, Lee J, Shin JH, Sung KC. Longitudinal changes in cardiovascular-kidney-metabolic syndrome stages and their impact on outcomes: a nationwide cohort study. J Clin Med 2025 Jun 1;14(11):3888. [doi: 10.3390/jcm14113888] [Medline: 40507650]
92. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). J Am Coll Cardiol 2018 Oct 30;72(18):2231-2264. [doi: 10.1016/j.jacc.2018.08.1038] [Medline: 30153967]
93. Shabtai R, Gatuz MV, Folman A, et al. Impact of cardiovascular-kidney-metabolic syndrome staging on myocardial infarction outcomes: a retrospective analysis of 2.7 million patients. Diseases 2025 Mar 27;13(4):97. [doi: 10.3390/diseases13040097] [Medline: 40277808]
94. Virani SS, Newby LK, Arnold SV, et al. 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA guideline for the management of patients with chronic coronary disease: a report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. Circulation 2023 Aug 29;148(9):e9-e119. [doi: 10.1161/CIR.0000000000001168] [Medline: 37471501]
95. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 2022 May 3;79(17):e263-e421. [doi: 10.1016/j.jacc.2021.12.012] [Medline: 35379503]
96. Joglar JA, Chung MK, Armbruster AL, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2024 Jan 2;149(1):e1-e156. [doi: 10.1161/CIR.0000000000001193] [Medline: 38033089]

XSL•FO
RenderX

97. Criqui MH, Matsushita K, Aboyans V, et al. Lower extremity peripheral artery disease: contemporary epidemiology, management gaps, and future directions: a scientific statement from the American Heart Association. Circulation 2021 Aug 31;144(9):e171-e191. [doi: 10.1161/CIR.0000000000001005] [Medline: 34315230]
98. Gold R, Larson AE, Sperl-Hillen JM, et al. Effect of clinical decision support at community health centers on the risk of cardiovascular disease: a cluster randomized clinical trial. JAMA Netw Open 2022 Feb 1;5(2):e2146519. [doi: 10.1001/jamanetworkopen.2021.46519] [Medline: 35119463]
99. Sperl-Hillen JM, Crain AL, Margolis KL, et al. Clinical decision support directed to primary care patients and providers reduces cardiovascular risk: a randomized trial. J Am Med Inform Assoc 2018 Sep 1;25(9):1137-1146. [doi: 10.1093/jamia/ocy085] [Medline: 29982627]
100. Columbo JA, Daya N, Colantonio LD, et al. Derivation and validation of ICD-10 codes for identifying incident stroke. JAMA Neurol 2024 Aug 1;81(8):875-881. [doi: 10.1001/jamaneurol.2024.2044] [Medline: 38949838]
101. Kuang A, Xu C, Southern DA, Sandhu N, Quan H. Validated administrative data based ICD-10 algorithms for chronic conditions: a systematic review. J Epidemiol Popul Health 2024 Aug;72(4):202744. [doi: 10.1016/j.jeph.2024.202744] [Medline: 38971056]
102. Claudel SE, Verma A. Cardiovascular-kidney-metabolic syndrome: a step toward multidisciplinary and inclusive care. Cell Metab 2023 Dec 5;35(12):2104-2106. [doi: 10.1016/j.cmet.2023.10.015] [Medline: 38056429]
103. Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. Ann Intern Med 2020 Jun 2;172(11 Suppl):S79-S84. [doi: 10.7326/M19-0873] [Medline: 32479175]
104. Fleet JL, Dixon SN, Shariff SZ, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. BMC Nephrol 2013 Apr 5;14(81):81. [doi: 10.1186/1471-2369-14-81] [Medline: 23560464]
105. Grams ME, Plantinga LC, Hedgeman E, et al. Validation of CKD and related conditions in existing data sets: a systematic review. Am J Kidney Dis 2011 Jan;57(1):44-54. [doi: 10.1053/j.ajkd.2010.05.013] [Medline: 20692079]
106. Jalal K, Anand EJ, Venuto R, Eberle J, Arora P. Can billing codes accurately identify rapidly progressing stage 3 and stage 4 chronic kidney disease patients: a diagnostic test study. BMC Nephrol 2019 Jul 12;20(1):260. [doi: 10.1186/s12882-019-1429-4] [Medline: 31299918]
107. Fernández-Llaneza D, Hilbrands LB, Vogt L, Engberink R, Klopotowska JE, LEAPfROG Consortium. Identifying a cohort of hospitalized chronic kidney disease patients using electronic health records: lessons learnt and implications for future research and clinical practice guidelines. Clin Kidney J 2025 Apr;18(4):sfaf073. [doi: 10.1093/ckj/sfaf073] [Medline: 40226368]
108. Nelson SJ, Yin Y, Trujillo Rivera EA, et al. Are ICD codes reliable for observational studies? Assessing coding consistency for data quality. Digit Health 2024;10:20552076241297056. [doi: 10.1177/20552076241297056] [Medline: 39493629]
109. Wiley LK, Rasmussen LV, Levinson RT, et al. PhenoFit: a framework for determining computable phenotyping algorithm fitness for purpose and reuse. J Am Med Inform Assoc 2026 Feb 1;33(2):536-542. [doi: 10.1093/jamia/ocaf195] [Medline: 41223026]
110. Wei WQ, Rowley R, Wood A, MacArthur J, Embi PJ, Denaxas S. Improving reporting standards for phenotyping algorithm in biomedical research: 5 fundamental dimensions. J Am Med Inform Assoc 2024 Apr 3;31(4):1036-1041. [doi: 10.1093/jamia/ocae005] [Medline: 38269642]
111. Monestime JP, Mayer RW, Blackwood A. Analyzing the ICD-10-CM transition and post-implementation stages: a public health institution case study. Perspect Health Inf Manag 2019;16(Spring):1a. [Medline: 31019430]
112. Fung KW, Xu J, Bodenreider O. The new international classification of diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM. J Am Med Inform Assoc 2020 May 1;27(5):738-746. [doi: 10.1093/jamia/ocaa030] [Medline: 32364236]

Abbreviations

AF: atrial fibrillation
AHA: American Heart Association
AMI: acute myocardial infarction
ASCVD: atherosclerotic cardiovascular disease
BP: blood pressure
CKD: chronic kidney disease
CKM: cardiovascular-kidney-metabolic
CT: computed tomography
CVD: cardiovascular disease
eGFR: estimated glomerular filtration rate
EHR: electronic health record
ESRD: end-stage renal disease
FY2026: Fiscal Year 2026

XSL•FO
RenderX

HF: heart failure
***ICD*:** *International Classification of Diseases*
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
***ICD-10-CM*:** *International Classification of Diseases, Tenth Revision, Clinical Modification*
KDIGO: Kidney Disease: Improving Global Outcomes
LVH: left ventricular hypertrophy
MI: myocardial infarction
NAFLD: nonalcoholic fatty liver disease
NASH: nonalcoholic steatohepatitis
OSA: obstructive sleep apnea
PAD: peripheral artery disease
PCOS: polycystic ovary syndrome
SHPT: secondary hyperparathyroidism
T2DM: type 2 diabetes mellitus

Edited by S Li; submitted 20.Jan.2026; peer-reviewed by A Padiyar, T Zhang, Y Cao; revised version received 03.Apr.2026; accepted 12.Apr.2026; published 01.Jun.2026.

Please cite as:
Zhao M, Zhang J, Wang H, Xu C, Li J, Li C, He X, Zheng X
Cardiovascular-Kidney-Metabolic Syndrome: Development of an ICD-10-CM Coding Framework
JMIR Diabetes 2026;11:e91827
URL: https://diabetes.jmir.org/2026/1/e91827
doi:10.2196/91827

© Minzhe Zhao, Junhao Zhang, Huiyun Wang, Chongjun Xu, Jiahui Li, Chenghua Li, Xuemei He, Xueying Zheng. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 1.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

User-Centered Development of a Digital Health Service for Diabetic Foot Ulcer Risk Stratification: Usability Study

Ulla Hellstrand Tang[1,2], PhD; Shivani Ravichandran[3], MSc; Stefan Candefjord[3], PhD; Dipu T Sathyapalan[4], PhD; Vivek Lakshmanan[5], PhD

[1]Department of Prosthetics and Orthotics at the Department for Biomedical Engineering and Medical Physics, Sahlgrenska University Hospital, Falkenbergsgatan 3, Gothenburg, Sweden

[2]Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[3]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

[4]Division of Infectious Diseases, Amrita Institute of Medical Science and Research Centre, Department of General Medicine, Amrita Vishwa Vidyapeetham, Kochi, Kerala, India

[5]Department of Podiatry, Amrita Institute of Medical Science and Research Centre, Amrita Vishwa Vidyapeetham, Kochi, Kerala, India

Corresponding Author:
Ulla Hellstrand Tang, PhD
Department of Prosthetics and Orthotics at the Department for Biomedical Engineering and Medical Physics, Sahlgrenska University Hospital, Falkenbergsgatan 3, Gothenburg, Sweden

Abstract

Background: Globally, 537 million persons live with diabetes, and a lifetime risk of up to 34% of developing diabetic foot ulcers (DFUs) necessitates strengthened preventive initiatives.

Objective: The study aimed to develop and evaluate a clinical decision support system (CDSS) to be used by health care professionals in foot assessment and risk stratification as a base for prevention.

Methods: Based on principles of human-computer interaction, the CDSS was developed for DFU risk assessment. Users, health care professionals from Region Västra Götaland in Sweden, evaluated the functions regarding effectiveness, efficiency, and satisfaction using a mixed methods usability testing approach. Expectations and experiences of using the CDSS were evaluated with the System Usability Scale (SUS).

Results: A total of 9 participants participated. User expectations of the CDSS, measured by SUS, averaged 77.2 (SD 14.6). Posttest SUS scores were 68.9 (SD 14.3), with a mean difference of 8.3 (P=.07), a nonsignificant reduction of usability after testing. The effectiveness of the CDSS in supporting users to complete 9 clinical tasks showed that for 7 (78%) tasks, at least 5 (56%) testers successfully achieved the intended goals. Tasks involving the identification of ingrown toenails and the confirmation of foot status, including risk stratification for the patient, were completed by fewer testers. Efficiency, measured as mean task completion time, ranged from 7 seconds to 9 minutes 20 seconds, and qualitative feedback informed recommendations for further system refinement. Users reported that a structured CDSS has the potential to support more equitable, consistent, and person-centered DFU prevention within a digital health service.

Conclusions: A digital health service for DFU risk stratification was developed based on national and international guidelines. Although the users' expectations of the usability were higher compared to how they experienced the CDSS, the SUS test was near a threshold of 70, indicating that the system being tested was above average in usability. Further development and validation, both nationally and internationally, with continued attention to users' needs and contextual factors, are recommended.

Trial Registration: ClinicalTrials.gov NCT05692778; https://clinicaltrials.gov/study/NCT05692778

(JMIR Diabetes 2026;11:e83287) doi:10.2196/83287

KEYWORDS

diabetic foot; diabetes mellitus; user-centered design; clinical decision support system; digital health

Introduction

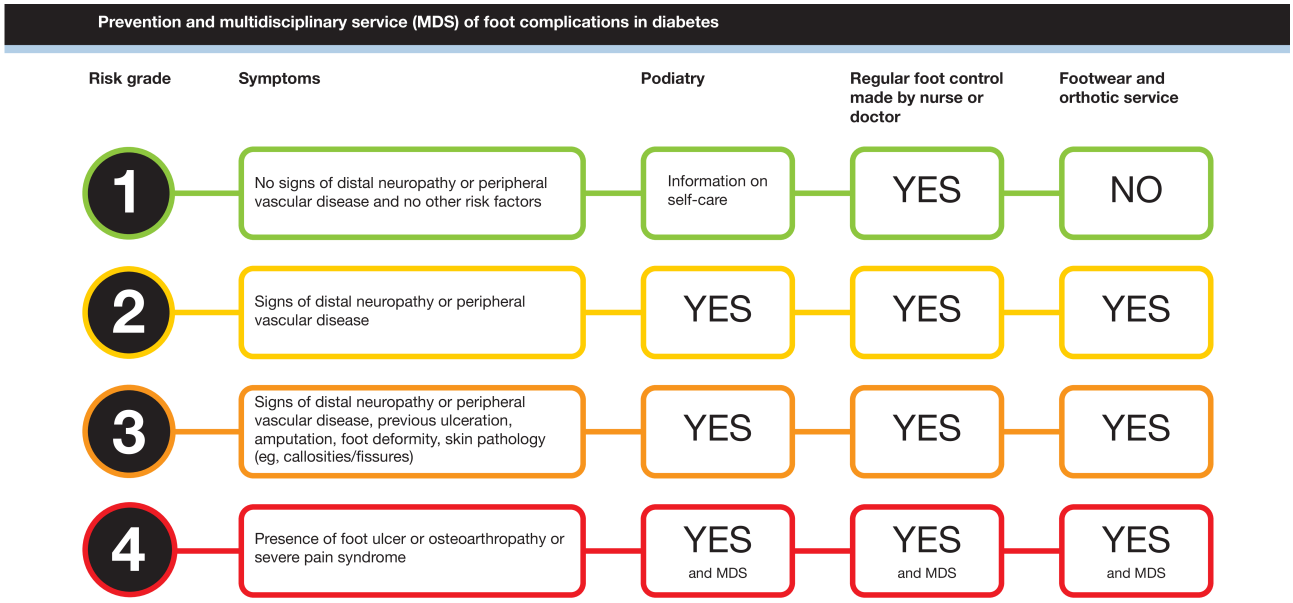
Background

Globally, 589 million persons live with diabetes [1], with a lifetime risk of developing diabetic foot ulcers (DFUs) ranging from 19% to 34% [2]. Out of 10.6 million people in Sweden, 600,000 persons had diabetes in 2024, according to the Swedish National Diabetes Register (NDR) [3]. In India, an increase from 90 million people with diabetes in 2024 to 157 million in 2050 is estimated [1]. DFUs are a serious and growing global health challenge. When left untreated or not promptly addressed,

XSL•FO
RenderX

DFUs can lead to severe complications, including lower-extremity amputation, significantly affecting patients' quality of life and increasing health care costs [4,5]. The 5-year mortality rate following DFU is 30% due to complications related to diabetes and comorbidities [6]. Worldwide, 18.6 million people live with DFU, with vascular, neurological, and biomechanical components [2]. The International Working Group on the Diabetic Foot identifies the following key risk factors for DFU development: peripheral neuropathy, peripheral vascular disease, skin pathologies, foot deformities, and a history of previous DFU or amputations [7] (Figure 1). These risk factors need to be identified regularly in each person living with diabetes. Despite well-established guidelines recommending regular foot examinations and risk assessments, a significant proportion of at-risk patients do not undergo structured screening for DFU risk in Swedish clinical practice. For instance, data from NDR indicate that 25% of patients in Sweden do not undergo the recommended foot examinations [3]. Timely interventions, such as podiatry and therapeutic footwear, may be omitted. Early prevention is recommended to hinder DFU progression and reduce the risk of amputation, and rapid referral to multidisciplinary care teams is recommended for patients diagnosed with DFUs [7].

Figure 1. Schema of the risk stratification as described in Swedish national clinical guidelines [8].



The Role of Digital Health Services in the Prevention of Foot Ulcers

The accelerated digitalization of health care has prompted efforts to integrate clinical decision support systems (CDSSs) [9] that allow health care professionals (HCPs) to register patient data, store it securely, and use it for evaluation and audit, and potentially for artificial intelligence (AI)–driven outcome prediction. However, despite efforts to develop digital health solutions, promoting foot health for patients with diabetes, no integrated system is currently in use in Sweden or elsewhere (as of 2025). On a global scale, Kabir and Ashmed [10], in their review, found that existing software for DFU management lacks sufficient evidence-based reliability, emphasizing the need for systematic evaluation. Their review of 170 applications identified major shortcomings, such as inadequate assessment tools, poor integration of best practices, and limited use of AI-driven enhancements. To ensure that these applications meet health care standards, a robust evaluation framework, such as human-computer interaction (HCI), is crucial, ultimately enhancing DFU management, improving patient outcomes, and reducing health care costs [10].

Theoretical Framework: HCI

The field of HCI has evolved beyond its initial focus on computer-centric usability to encompass diverse technologies, including CDSS. HCI research aims to optimize interactions between users and digital tools. Bevan and Harker [11] contributed to improving human interaction design in health care by refining the ISO 9241 - 11 usability standard [12], emphasizing a more comprehensive understanding of usability beyond efficiency, effectiveness, and satisfaction. User-centered design is an iterative development methodology that prioritizes user needs, tasks, and environmental factors throughout the design and implementation processes. This approach ensures that digital tools align with end users' requirements, thereby enhancing adoption and clinical use. Furthermore, evaluating the usability of the system is crucial in order to ensure that the system fulfills the needs and performs effectively in clinical settings [13].

Grundgeiger et al [14] advocated for incorporating user experience (UX) principles in safety-critical health care environments. They emphasized that beyond usability, factors such as emotional response, trust, and cognitive workload significantly impact HCPs' interactions with technology. UX refers to the overall impact of a user's interaction with a system,

XSL•FO
RenderX

spanning initial impressions, long-term engagement, and emotional responses. Usability, a core component of UX, refers to the effectiveness, efficiency, and satisfaction with which users achieve specified goals in a given context. The concept of "UX" by Norman et al [15] highlights the importance of a holistic interaction design beyond interface usability.

A limited number of applications, such as CONNECTPlus [16] and D-Foot [17-19], have been specifically developed for HCPs to support systematic identification of DFU risk and structured foot assessment. Both D-Foot and CONNECTPlus aim to facilitate risk stratification and support clinical decision-making by providing structured assessment tools and guidance.

D-Foot, developed within Region Västra Götaland (VGR) in western Sweden, was designed for use by certified prosthetists and orthotists. The system is based on a structured foot examination aligned with recommendations from the Swedish Association of Local Authorities and Regions [8,20].

CONNECTPlus offers features such as risk-specific education, self-management support, and remote monitoring intended to empower patients and potentially reduce health care burden; however, its direct impact on clinical outcomes has not yet been established.

Prior to this study (in 2020), the principal investigator (UT) led a regional development group in the VGR that, based on national guidelines [20], created a workflow for a foot examination in paper format to facilitate risk stratification for patients with diabetes (Multimedia Appendix 1). The face validity of the foot examination was secured by HCPs regionally and nationally during 2 workshops. The foot examination in paper format, with its structured foot assessment, has previously been explored regarding how HCPs experienced its use during foot assessments of patients with diabetes [17]. The HCPs found that the foot examination simplified examinations, highlighted differences in current standard management, and had the potential to standardize care and achieve good, equal, and person-centered care [17]. While some inconsistencies in risk categorization were noted, the anticipated digital version was expected to enhance documentation, accessibility, and adherence to guidelines.

Aim and Research Question

Despite digital advancements, no CDSS is incorporated in a digital health service aimed to be used by HCPs, such as nurses, podiatrists, and doctors, for the early detection of risk factors that precede the development of DFUs. To address this gap, this study presents the design and evaluation of a digital health service including a CDSS for managing structured foot examinations.

The specific research questions were as follows:

1. How should a CDSS be designed to effectively capture and support the workflow in diabetic foot care?
2. Does the CDSS function as intended in meeting user expectations and supporting HCPs in real-world settings?
3. How does usability with pretest and posttest surveys contribute to refining the design of a CDSS?

Methods

Design and Setting

The study was part of a larger research initiative aimed at optimizing the prevention and care of patients with diabetes at risk of developing DFUs. The study was conducted in collaboration with VGR, the public health care provider for the second largest region in Sweden, serving approximately 1.8 million inhabitants. The study was carried out in 2 health care settings within VGR: Skaraborg Hospital in Skövde and Sahlgrenska University Hospital in Gothenburg (Multimedia Appendix 2). Usability testing sessions were conducted at these hospital sites during 2022. Each session lasted approximately 2 hours and was designed to allow in-depth observation and interaction with the CDSS in a controlled clinical environment. In addition, at a later stage, researchers from Amrita University in India were involved as a reference group.

The methodology comprised three phases—(1) design, (2) testing, and (3) iterative refinement—following established usability testing principles described by Dumas and Redish [21].

Design Phase

In the design phase, the following steps were taken:

1. An innovation team, consisting of a clinical researcher, a researcher in health informatics, a master's student in biomedical engineering, and a researcher in biomedical engineering, formulated the strategy for the development of the digital health service.
2. The requirements for the digital health service were defined based on the foot examination that HCPs had created (Multimedia Appendix 1), a literature review, patients' needs as expressed in the national clinical guidelines [8], and a review of existing diabetic foot care workflows in the VGR.
3. The key steps in diabetic foot care were identified and integrated into the design. Personas were created (Multimedia Appendix 3). Personas are useful in HCI and are fictional yet research-based representations of specific user groups, helping designers [22]. They guide design decisions by encapsulating user characteristics, behaviors, and goals, ensuring usability and accessibility in the development of interactive systems [23,24].
4. The software Lucidchart [25] (Multimedia Appendix 4), a web-based diagramming application, was used to create the conceptual design by visualizing the workflow anticipated in the digital health service during foot examination. The workflow for HCPs included useful features such as an education section for enhanced learning during use, an examination section, and a documentation section.
5. In the development of the digital health service, the software Figma (Dylan Field and Evan Wallace), a collaborative web application useful in interface design, was used (Multimedia Appendix 5) [26].

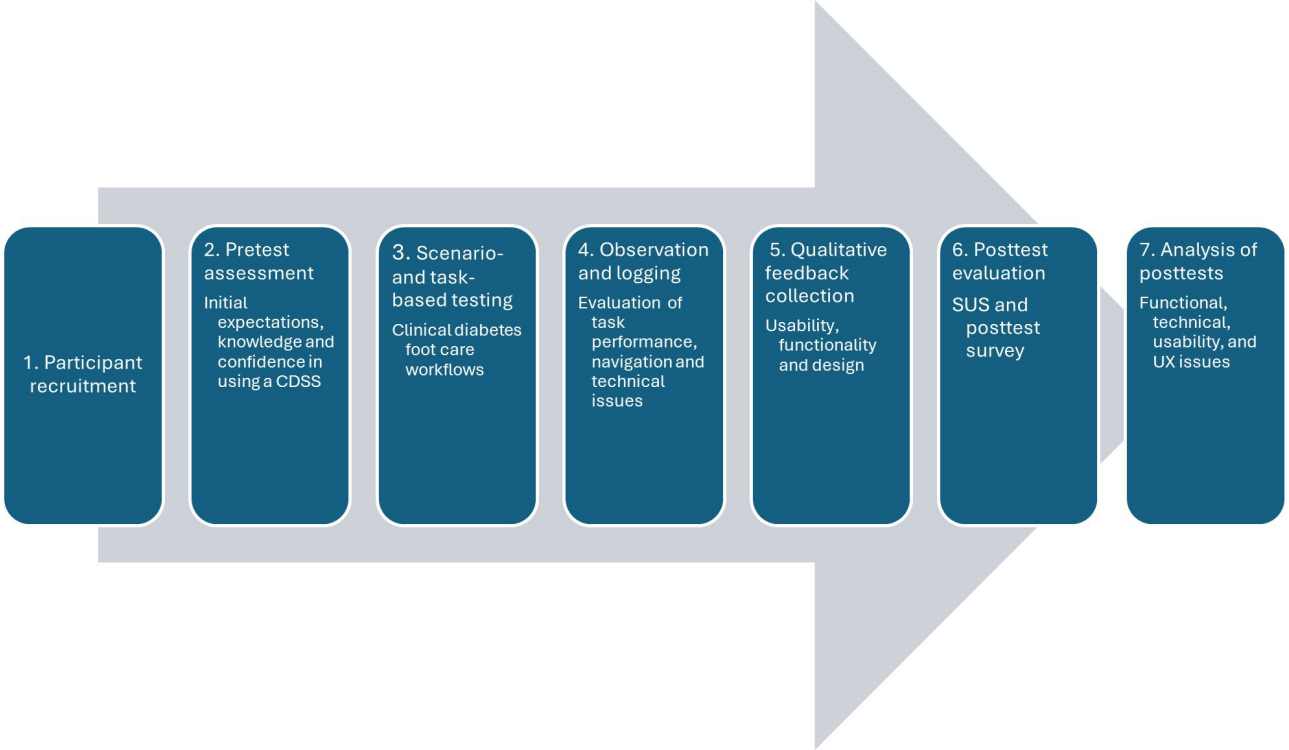
The Test Phase

The usability testing followed a structured, sequential process (Figure 2). First, users' initial expectations, knowledge, and

XSL•FO
RenderX

confidence in using the CDSS were assessed using a pretest questionnaire (Multimedia Appendix 6). This was followed by scenario- and task-based testing, during which HCPs performed tasks reflecting routine diabetic foot care workflows. User interactions were observed to identify technical issues and assess whether functional requirements were met.

Figure 2. The test phases. CDSS: clinical decision support system; SUS: System Usability Scale; UX: user experience.



Qualitative feedback was then collected regarding usability, functionality, and design. Usability was further evaluated using pretest and posttest surveys, including the System Usability Scale (SUS), to assess user expectations and satisfaction [27-30]. Finally, posttest surveys (Multimedia Appendix 7) were used to assess users' perceptions of how well the CDSS met their needs and expectations, as well as to identify issues related to ease of use, navigation, and overall UX.

Participants

Participants were HCPs employed within VGR and were recruited using a purposive sampling approach. Inclusion criteria were that participants had clinical experience working with patients with diabetes at risk of developing DFUs. Eligible HCPs were systematically contacted by a member of the research team, informed about the study, and invited to participate in the usability testing.

A total of 9 HCPs participated in the study, including 3 nurses, 4 podiatrists, 1 physiotherapist, and 1 certified prosthetist and orthotist. Participants' ages ranged from 37 to 68 years, and their clinical experience in diabetes care ranged from 1 to 40 years. To support detailed observation and interaction, each usability testing session included a maximum of 3 participants. The number of participants was judged to be sufficient, given the diversity of professions among testers and considering the cost of testing against the value of the information gained [31].

Ethical Considerations

The study was approved by the Swedish Ethical Review Authority (register number 2020 - 02715) and was conducted according to the ethical principles described in the Declaration of Helsinki, following the Code of Ethics of the World Medical Association for experiments involving humans [32]. The participants received oral and written information about the study, including information on their right to withdraw from the study at any time without explanation. Participants were included after providing written informed consent, and all participants agreed to be photographed. Written informed consent was also obtained for the publication of any potentially identifiable images or data included in this article. There is always a risk of intrusion of integrity when personal experiences are shared with others. However, the advantages for future patients were judged to balance the possible inconvenience that participants might experience, for example, the time they spent on travel and participation in the study. The participants were ensured confidentiality and were informed that any concerns could be clarified by contacting the research team. No such concerns were expressed. The ClincalTrials.gov ID is NCT05692778.

Written informed consent was obtained from all individuals depicted in figures for publication of their images in this open access article. Privacy measures were implemented to protect participants. No personally identifying information (eg, names or other identifiers) is included in the image or accompanying text. The images are used solely for illustrative purposes within the context of the study. Participant data were de-identified prior to analysis, and no directly identifiable information (eg, names or personal identifiers) was included in the dataset. No such concerns were expressed. No compensation was provided to the individuals.

XSL•FO
RenderX

Workshop for Usability Testing

The usability test was held in the Skaraborg region, at Skövde Hospital and in Gothenburg during 2022. First, during the workshop, the study participants filled in a survey capturing demographic data, professional background, and experience of using digital tools in assessing foot status in patients with diabetes. Participants answered a mix of closed-ended (yes/no, multiple choice) and open-ended questions with pretest surveys (Multimedia Appendix 6). Furthermore, the HCPs' prior experiences and their expectations of working with digital tools were registered. During the tests, the participants performed simulated user tasks, and the usability was evaluated with the following measures: (1) the time it took to perform the tasks and (2) registration in a think-aloud protocol, meaning that users' thoughts were registered while performing tasks [33].

None of the 9 testers had, at the time of the workshop, previously used digital tools for assessing foot status in patients with diabetes. Six testers had not previously used digital tools for assessing patients' health either. However, 2 testers had registered health status in the NDR [3], and 1 tester had documented foot status of the patient with a photo stored in software, Piscara [34]. Eight of the 9 testers took photos of the patient's foot during examination in clinical practice but without registering where the photos were stored. Five of the testers preferred to use iPad [35], 3 testers used Android systems (eg, Samsung, Xiaomi, OnePlus, and Motorola), and 1 tester did not know which system they used.

The usability test was introduced by 2 persons from the study team (SR and UT). The observers (n=2) made observations during the tests and registered comments from the testers, according to the think-aloud method [33]. The test sessions were recorded to be able to recall the comments made. The rooms used for the usability test were arranged so that testers had all the necessary equipment and information at hand (Figure 3), for example, a file, in paper format, was assigned to each of the testers with all instructions and the pretest and posttest surveys.

The English language was used, since the study was a collaboration with English-speaking researchers at Chalmers University of Technology and Amrita Vishwa Vidyapeetham, Kochi, in India. During the tests, the testers were free to ask questions in Swedish for clarification, and think-aloud comments were made both in English and in Swedish.

Figure 3. Usability testing of the digital health service. The testers were provided with tablets and documents in paper format with information about the usability test and the pretest and posttest surveys, and instructions on how to perform the usability tasks.

XSL•FO
RenderX

Statistics

The study used a mixed methods design, combining qualitative and quantitative methods [36]. Quantitative data were analyzed using descriptive statistics and hypothesis testing. Participant characteristics were summarized using counts and percentages for categorical variables and means (SDs) for continuous variables. The SUS consists of 10 items rated on a 5-point Likert scale. Although individual items are ordinal, the SUS total score is calculated according to standard procedures to yield a composite score ranging from 0 to 100. Consistent with established practice, the aggregated SUS score was treated as continuous data and analyzed using parametric statistical methods, including paired Student *t* tests, to compare pretest and posttest scores [37], with a significance threshold set at $P<.05$. Task efficiency was assessed by measuring the time required to complete each predefined task and is reported as mean values with SDs. Task effectiveness was evaluated by calculating the proportion of participants who successfully completed each task.

Qualitative data from think-aloud protocols, observational notes, and open-ended survey responses were analyzed using an inductive content analysis approach supported by AI-assisted categorization. The AI tool was used to facilitate initial identification and grouping of meaning units in a data-driven manner. All preliminary codes and categories were subsequently reviewed, refined, and validated by the research team to ensure they accurately reflected the original data and study context. Final categorizations and interpretations were made by the researchers.

The study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) eHealth checklist (Checklist 1).

Results

User expectations of the digital health service, measured with the SUS pretest, were 77.2 (SD 14.6), and the posttest SUS scores were 68.9 (SD 14.3), with a mean difference of 8.3 ($P=.07$), indicating a nonsignificant reduction in perceived usability after use. The effectiveness of the digital health services in supporting users to complete 9 specified clinical tasks showed that for 7 (78%) tasks, at least 5 (56%) testers successfully achieved the intended goals (Table 1). Tasks involving the identification of ingrown toenails and confirmation of foot status and risk stratification for a fictitious patient were completed by fewer testers. Participants expressed that a structured CDSS has the potential to contribute to more equitable and person-centered DFU prevention. Efficiency, measured as mean task completion time, ranged from 7 seconds to 9 minutes 20 seconds. The final task was to discuss the care plan with the fictitious patient, based on the risk stratification and foot status. One of the testers suggested that the foot assessment and the risk grade should be printed as a base for further discussion regarding self-care and care. In addition, it was suggested that the need for podiatry and the use of appropriate shoes should be discussed with the patients. Furthermore, it was suggested that the CDSS should automatically allow sending digital referrals for podiatry, to the Department of Prosthetics and Orthotics, and to specialist care.

XSL•FO
RenderX

Table . Effectiveness and efficiency of the usability tests (n=9).[a]

Task and specification of goals to reach	Effectiveness (goal reached)			Efficiency, mean (SD), mm:ss
	Yes, n (%)	No, n (%)	Missing, n (%)	
Warm-up task				
To enter the app	N/A[b]	N/A	N/A	05:15 (05:15)
Warm-up task				
To select your patient	N/A	N/A	N/A	00:23 (00:16)
Examine for pressure area				
Did the testers save Yes (pressure) for the left and right foot?	6 (67)	0 (0)	3 (33)	Missing
Did the testers examine both feet?	7 (78)	0 (0)	2 (22)	Missing
Examine for ingrown nails				
Did the testers save Yes (ingrown nails) for the right foot?	4 (44)	1 (11)	4 (44)	02:06 (01:13)
Examine for hallux valgus				
Did the testers click on the “Learn about” button?	5 (56)	3 (33)	1 (11)	00:37 (00:31)
Did the testers save Yes (hallux valgus) for the right foot?	7 (78)	1 (11)	1 (11)	__[c]
Ask the fictitious patient if he has peripheral neuropathy				
Did the testers save “Loss of sensation” for the left and the right foot?	6 (67)	1 (11)	2 (22)	02:39 (02:39)
Examine the fictive patient to check if he has peripheral neuropathy				
Did the testers click on the “Learn about” button?	7 (78)	1 (11)	1 (11)	03:38 (02:42)
Did the testers save “Loss” for the left and the right foot	7 (78)	0 (0)	2 (22)	—
Look at the summary and confirm that results are “OK”				
Look at the summary and confirm that results are “OK”	0 (0)	0 (0)	9 (100)	00:48 (01:12)
Discuss the care plan with the fictitious patient				
Did the testers show the results to the fictitious patient?	6 (67)	0 (0)	3 (33)	02:04 (01:48)
What discussion did the nurse/tester have with the fictive patient regarding the care plan?	1 (11)	0 (0)	8 (89)	—

[a]This table summarizes the results of the usability testing in terms of effectiveness, efficiency, and task completion metrics. Nine main tasks were included. In the column “Specification“ column, details about each task are presented. Efficiency, meaning the time spent to execute the tasks, is

XSL•FO
RenderX

presented in the "Efficiency" column.

[b]N/A: not applicable.

[c]Not available.

Several useful suggestions for refining the design of the CDSS were recorded, based on the think-aloud protocols. The following themes were identified: (1) referrals and communication, (2) foot examination features, (3) user interface and navigation, (4) training and usability, (5) technical issues, (6) workflow and click analysis, (7) content feedback, and (8) login and simplicity (Table 2).

Results from the posttest survey showed that all testers, except 1, would prefer to use a tablet in future digital foot examinations, with the rationale that the screen is larger compared with a smartphone screen. For 1 tester, the type of hardware did not matter. All testers would prefer an automatically generated risk scale in the final CDSS, commenting that an automatically generated risk scale could enhance clarity and flexibility and improve security regarding care. One comment was "Becomes clear and allows you to assess and inform about the result," and there were comments that the assessment made by the HCP of the overall need of each patient is necessary, assuring person-centered care. A total of 8 (89%) participants found it useful to register foot status for the left and the right foot separately and gave the rationale that it is of importance to address separate registration, thereby ensuring clarity and distinction. Instead of the current design, involving clicking forwards and backwards in the CDSS, 5 (56%) participants would prefer a scroll-down feature to navigate between questions. The structure of the CDSS was described as logical, clear, and good by 4 (44%) of the participants. A minor challenge noted was that it was "somewhat complicated, especially the language," as English was used instead of Swedish.

XSL•FO
RenderX

Table . Suggestions for improvements of the digital health service.

Theme	Feedback and comments
Referrals and communication	• Enable digital referrals to podiatry and prosthetics and orthotics based on the foot summary. • Provide an option to print or send a referral link to the patient's mobile. • Patients should be able to independently contact specialists (in person, by phone, or by email).
Foot examination features	• Ability to take and store photos of the feet (dorsal and plantar views). • Ensure image access for both patients and health care personnel.
User interface and navigation	• Include a "Back" button or icon on all question pages. • Allow zoom-in functionality for images. • Add a search function for the patient list. • Highlight "Yes/No" buttons upon selection for validation. • Consider removing the "Save" button if selection highlights clearly indicate that input has been registered. • Add visual confirmation (eg, click animation) to indicate when buttons are pressed. • Clarify whether all questions must be answered by the HCP[a].
Training and usability	• Provide user training prior to access. • Show tester feedback below each question. • Clarify whether illustrations (eg, ingrown nails) depict real cases or demos. • Address communication barriers between testers and observers.
Technical issues	• Device required firm pressure due to the screen protector. • Device performed better when handheld rather than lying flat. • "Learn about" button was overlooked during testing.
Workflow and click analysis	• Task 3 required 15 clicks to access the page and complete the task (tested 3 times). • Task 4 required 7 clicks before returning to "Back to inspection." • Suggest reducing the total number of clicks to simplify navigation.
Content feedback	• Ingrown toenail assessment was misinterpreted as evaluating thickened nails. • Illustrations need clearer contextual explanations.
Login and simplicity	• Request for a simpler login process.

[a]HCP: health care professional.

Discussion

Principal Findings

This study aimed to design and evaluate a digital health service to support HCPs in conducting structured foot assessments and risk stratification for DFUs. The principal findings indicate that the developed CDSS was usable and acceptable to HCPs across diverse professional backgrounds and varying levels of digital experience. Most participants were able to complete the majority of predefined clinical tasks, and the system demonstrated reasonable effectiveness and efficiency in supporting structured DFU assessment workflows. Although posttest usability ratings were slightly lower than pretest expectations, overall usability remained close to established acceptability thresholds. Importantly, participants perceived that a structured CDSS could support more consistent, equitable, and person-centered DFU prevention, aligning with the overarching objective of improving early detection and preventive care [38].

The usability and acceptance of the CDSS observed in this study are consistent with previous research indicating that structured foot assessment and digital decision support tools can assist HCPs in standardizing DFU assessments and improving adherence to clinical guidelines [38-42]. Similar to earlier usability studies of digital health interventions, participants were able to engage with the system despite limited prior experience with comparable tools, underscoring the importance of intuitive design and workflow alignment in early-stage implementations [43].

The observed decrease in SUS scores from pretest to posttest, although not statistically significant, reflects a pattern reported in other usability studies, where initial expectations tend to exceed postuse perceptions once users encounter practical constraints and design limitations [29,44]. This underscores the importance of formative usability testing in identifying mismatches between user expectations and real-world system performance. Tasks related to identifying specific foot

XSL•FO
RenderX

conditions and confirming final risk stratification were less consistently completed, suggesting that these cognitively demanding steps require clearer visualization and decision support. Previous studies on DFU risk assessment have similarly reported challenges in ensuring consistent interpretation of clinical findings, particularly when multiple risk factors must be integrated into a final classification [45]. Participants' emphasis on automated risk scoring and clearer differentiation between left and right foot assessments aligns with literature advocating for cognitive support features to reduce variability and enhance clinical decision-making [46]. Beyond usability, the perception that the CDSS could support person-centered and equitable care is in line with growing evidence that structured digital tools can improve communication, transparency, and shared understanding between HCPs and patients in chronic disease management [47].

Usability and Acceptance of the CDSS

The CDSS was generally well received by participating HCPs, representing varied professional backgrounds and clinical experience. Notably, although the majority of participants had little or no prior experience of using digital health services for diabetic foot assessments, they were able to engage with the current CDSS and complete most of the clinical tasks. Pretest SUS scores indicated high expectations (77.2, SD 14.6), while posttest scores were slightly lower (68.9, SD 14.3), though the difference was not statistically significant. This reduction suggests that while the tool was functional and promising, aspects of its usability require improvement to better meet user expectations.

Several usability challenges were identified through think-aloud protocols and posttest surveys, including different navigation options (eg, scroll-down vs page-by-page design), difficulties with English-language prompts, and requests for clearer supporting information. These findings underline the importance of an iterative user-centered design and the value of including real users in early testing stages to refine the tool for real-world use [12].

Effectiveness in Supporting Clinical Tasks and Workflow

The CDSS demonstrated reasonable effectiveness in supporting users through structured foot assessment workflows. Fifty-six percent of the participants completed 7 of 9 clinical tasks successfully, showing that the tool facilitated critical steps in DFU risk evaluation. However, tasks involving the identification of specific conditions (eg, ingrown toenails) and final risk stratification were less consistently completed, pointing to areas where the interface or decision logic could be enhanced. Participants emphasized the importance of an automatically generated risk score, separate documentation for the left and the right foot, and real-time feedback during the assessment. These suggestions align with the goal of supporting clinical reasoning and improving consistency in risk classification, which is vital for ensuring timely and accurate referrals and interventions. Importantly, the tool was perceived to promote equitable and person-centered care, a critical aspect of DFU prevention in both high-resource and resource-constrained settings [48].

Implications for Future Development and Implementation

Despite the growing digitalization in health care, few digital tools currently exist that are tailored for early DFU detection by HCPs. This study addressed a clear gap by developing a CDSS grounded in national guidelines, evidence-based practice, and real-world workflows. The digital health service has the potential to be scaled and localized for broader use, particularly given the global burden of diabetes and DFUs in Sweden and India, 2 countries represented in this project.

For successful implementation, future iterations of the CDSS should incorporate language adaptation, enhanced onboarding and training modules, and seamless integration with existing electronic health records (EHRs) or quality registers such as the Swedish NDR [3]. Furthermore, expanding tests to include patients and HCPs in clinical settings will be essential to validating the CDSS's effectiveness in reducing DFU incidence and improving patient outcomes.

To facilitate the usability test, the tests were conducted close to where the HCPs worked, in Skövde and Gothenburg, respectively. It was considered easier to perform the test on site where the HCPs worked, as they did not need to travel to a usability lab situated elsewhere. This arrangement allowed HCPs working in the city of Gothenburg in the Skaraborg region, that is, in the countryside, respectively, to contribute with their experiences of using the CDSS.

Methodological Considerations

A strength of this study was the integration of qualitative and quantitative data within a mixed methods framework. Findings from think-aloud sessions and user observations were triangulated with survey responses to identify convergent usability issues and inform iterative refinement of the CDSS. This integration enabled the translation of user feedback into concrete design improvements, including enhanced navigation functions and refinement of the output report to support patient communication and referral decisions. Another strength of the study was the diversity among the testers in professional backgrounds, experiences, and familiarity with technology.

A limitation of this study is the small sample size for the usability evaluation (n=9), which may be considered low from a research perspective. However, this number is consistent with established methodological guidance for formative usability testing in user-centered design, where the primary aim is to identify usability issues rather than to achieve statistical generalizability. Prior work demonstrates that most usability problems are detected with small samples and that returns diminish beyond approximately 5 to 10 participants [49-51]. Nevertheless, the findings should be interpreted as indicative rather than definitive, and additional usability issues may emerge with broader testing. Another limitation is that the sample may not fully capture the heterogeneity of intended end users, including variations in clinical settings, digital literacy, and demographic characteristics. In addition, a survey question regarding prior use of digital tools may have been misinterpreted, as participants did not consistently consider routine use of EHRs as use of a digital tool. The mixed use of

XSL•FO
RenderX

English and Swedish during usability testing also constituted a limitation, as it may have contributed to uncertainty during task performance and missing data. Furthermore, some visual elements and task formulations led to misinterpretation (eg, illustration of an ingrown toenail as a thickened toenail), indicating areas requiring refinement in future iterations of the CDSS.

Finally, a limitation is that implementation was not examined as an explicit object of inquiry using a formal implementation framework. Although organizational, contextual, and workflow-related issues emerged during usability testing, these were not systematically analyzed using an implementation science framework, such as the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability Clinical Assessment Tool [52]. Previous work from the western region of Sweden has highlighted complex challenges related to technical readiness and organizational preparedness for the development and implementation of CDSSs [53]. Consequently, determinants related to long-term adoption, scale-up, and sustainability in real-world clinical settings were not comprehensively assessed in the present study and should therefore be interpreted with caution.

Future Development

Future work should include summative usability evaluations and real-world implementation studies involving larger and more diverse user groups to assess generalizability, clinical effectiveness, and impact on patient outcomes. Further development must also align with regulatory requirements for medical devices, particularly the European Medical Device Regulation, including clinical validation, usability engineering, and risk management [54]. Beyond usability, future studies should evaluate the accuracy, reliability, and adaptability of automated risk stratification algorithms while ensuring that clinicians retain the ability to exercise clinical judgment. Integration with EHRs and quality registers, as well as structured onboarding and training for HCPs, will be essential for sustainable implementation. To support broader adoption, future research should explore cultural and regional adaptations of the CDSS, including use in resource-constrained settings, hardware preferences (eg, tablet vs smartphone), and workflow integration across different health care systems. Addressing these factors will be essential to realizing the full potential of digital decision support in advancing equitable and person-centered DFU prevention.

This study demonstrates the feasibility of developing a user-centered CDSS to facilitate structured foot assessment and DFU risk stratification in clinical practice. The findings indicate that the CDSS was usable and acceptable to HCPs with diverse backgrounds and levels of digital experience, and that it can support consistent, efficient, and person-centered preventive workflows. While further development, validation, and implementation-focused research are required, this work provides important formative evidence and design insights that can inform future digital decision support solutions aimed at improving early detection and prevention of diabetic foot complications.

Acknowledgments

The authors would like to express their sincere gratitude to all the participants of this study for their valuable time, insights, and engagement. Their contributions were essential to the success of this work. They also wish to thank the Innovation Platform in Region Västra Götaland for its support and collaboration throughout the project. Finally, they extend their heartfelt appreciation to the late Isabella Scandurra for her invaluable contributions and deep knowledge in human interaction design. Her expertise and dedication greatly enriched this study, and her influence continues to inspire our work. They are also most grateful for the graphical illustrations made by Pontus Andersson.

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing, and summarizing text.

The GAI tool used was ChatGPT (OpenAI; GPT-5.2).

Responsibility for the final manuscript lies entirely with the authors.

GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Declaration submitted by: UHT

Funding

The study was supported by the Innovation Platform in Region Västra Götaland, Gothenburg Diabetes Association, Stiftelsen Skobranschens Utvecklingsfond, Gunnar Holmgrens Minne, and Felix Neuberghs Foundation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: UHT, SR
Data curation: UHT, SR
Formal analysis: UHT, SR
Funding acquisition: UHT
Investigation: UHT, SR
Methodology: UHT, SR, SC
Project administration: UHT
Resources: UHT
Supervision: UHT, SC
Validation: UHT, SR, SC, DTS, VL
Visualization: UHT,
Writing – original draft: UHT, SR, SC, DTS, VL
Writing – review & editing: UHT, SR, SC, DTS, VL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structured foot examination.
[DOCX File, 28 KB - diabetes_v11i1e83287_app1.docx]

Multimedia Appendix 2
Skaraborg and Gothenburg in the Region Västra Götaland [55].
[DOCX File, 921 KB - diabetes_v11i1e83287_app2.docx]

Multimedia Appendix 3
Personas a fictive person living with diabetes and a foot ulcer.
[DOCX File, 455 KB - diabetes_v11i1e83287_app3.docx]

Multimedia Appendix 4
Workflow development for the digital health service using Lucidchart.
[JPG File, 70 KB - diabetes_v11i1e83287_app4.jpg]

Multimedia Appendix 5
Selected screenshots of clinical decision support system workflow evaluated in the workshop.
[DOCX File, 551 KB - diabetes_v11i1e83287_app5.docx]

Multimedia Appendix 6
Pretest survey.
[DOCX File, 27 KB - diabetes_v11i1e83287_app6.docx]

Multimedia Appendix 7
Posttest survey.
[DOCX File, 27 KB - diabetes_v11i1e83287_app7.docx]

Checklist 1
CONSORT-eHEALTH (V1.6.1) checklist.
[DOCX File, 69 KB - diabetes_v11i1e83287_app8.docx]

References

1. IDF diabetes atlas 2025. : International Diabetes Federation; 2025 URL: https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/ [accessed 2026-04-10]
2. Armstrong DG, Tan TW, Boulton AJM, Bus SA. Diabetic foot ulcers: a review. JAMA 2023 Jul 3;330(1):62-75. [doi: 10.1001/jama.2023.10578] [Medline: 37395769]
3. Nationella Diabetesregistret [Website in Swedish]. 2025. URL: https://ndr.registercentrum.se [accessed 2026-04-13]

XSL•FO
RenderX

4. Edmonds M, Manu C, Vas P. The current burden of diabetic foot disease. J Clin Orthop Trauma 2021 Jun;17:88-93. [doi: 10.1016/j.jcot.2021.01.017] [Medline: 33680841]
5. McDermott K, Fang M, Boulton AJM, Selvin E, Hicks CW. Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers. Diabetes Care 2023 Jan 1;46(1):209-221. [doi: 10.2337/dci22-0043] [Medline: 36548709]
6. Armstrong DG, Swerdlow MA, Armstrong AA, Conte MS, Padula WV, Bus SA. Five year mortality and direct costs of care for people with diabetic foot complications are comparable to cancer. J Foot Ankle Res 2020 Mar 24;13(1):16. [doi: 10.1186/s13047-020-00383-2] [Medline: 32209136]
7. Schaper NC, van Netten JJ, Apelqvist J, et al. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3657. [doi: 10.1002/dmrr.3657] [Medline: 37243927]
8. Personcentrerat och sammanhållet vårdförlopp diabetes med hög risk för fotsår [Report in Swedish]. : Sveriges Kommuner och Regioner; 2022 URL: https://d2flujgsl7escs.cloudfront.net/external/Vardforlopp_Diabetes.pdf [accessed 2026-04-06]
9. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3(1):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
10. Kabir A, Ahmed S. Future of foot care: mobile apps in diabetic ulcer management. Presented at: Australian National Pedorthic Conference 2024; Sep 12-14, 2024 URL: https://researchoutput.csu.edu.au/en/publications/future-of-foot-care-mobile-apps-in-diabetic-ulcer-management/ [accessed 2026-04-06]
11. Bevan N, Harker S. ISO 9241-11 revised: what have we learnt about usability since 1998? Presented at: Human-Computer Interaction: Design and Evaluation; Aug 2-7, 2015. [doi: 10.1007/978-3-319-20901-2_13]
12. ISO 9241-210:2019 ergonomics of human-system interaction—part 210: human-centred design for interactive systems. : International Organization for Standardization (ISO); 2019 URL: https://cdn.standards.iteh.ai/samples/77520/8cac787a9e1549e1a7ffa0171dfa33e0/ISO-9241-210-2019.pdf [accessed 2026-04-10]
13. Innovating for People: Handbook of Human-Centered Design Methods: Luma Institute: Luma Institute; 2012. URL: https://www.luma-institute.com/wp-content/uploads/2021/02/22Innovating-for-People22-Looking.pdf [accessed 2026-04-06]
14. Grundgeiger T, Hurtienne J, Happel O. Why and how to approach user experience in safety-critical domains: the example of health care. Hum Factors 2021 Aug;63(5):821-832. [doi: 10.1177/0018720819887575] [Medline: 31914323]
15. Norman DA, Miller J, Henderson A. What you see, some of what's in the future, and how we go about doing it: HI at apple computer. Presented at: CHI '95: Conference companion on Human factors in computing systems; May 7-11, 1995 URL: https://www.researchgate.net/publication/202165701_What_You_See_Some_of_What's_in_the_Future_And_How_We_Go_About_Doing_It_HI_at_Apple_Computer [accessed 2026-04-06]
16. CONNECTPlus: reducing the risk of diabetic foot ulceration. Health and Care Innovations Limited. 2025. URL: https://www.hci.digital/connect-plus/diabetic-foot [accessed 2026-04-06]
17. Andersson S, Scandurra I, Nyström U, Varemo M, Tang UH. Experiences of a novel structured foot examination form for patients with diabetes from the perspective of health care professionals: qualitative study. JMIR Nurs 2023 Jul 18;6:e45501. [doi: 10.2196/45501] [Medline: 37463012]
18. Hellstrand Tang U, Tranberg R, Sundberg L, Scandurra I. How do patients and healthcare professionals experience foot examinations in diabetes care? - A randomised controlled study of digital foot examinations versus traditional foot examinations. BMC Health Serv Res 2024 Nov 12;24(1):1387. [doi: 10.1186/s12913-024-11674-w] [Medline: 39533310]
19. Hellstrand Tang U, Tranberg R, Zügner R, et al. The D-Foot, for prosthetists and orthotists, a new eHealth tool useful in useful in risk classification and foot assessment in diabetes. Foot Ankle Online J 2017;10 [FREE Full text]
20. Östensson CG, Johansson K, Insulander L, Jonsson LV, Löndahl M, Sergu-Bogdan C, editors. Fotundersökning vid diabetes-nationellt vårdprogram för prevention av fotkomplikationer vid diabetes [Report in Swedish]. : Sveriges kommuner och regioner; 2018 URL: https://swenurse.se/download/18.32612fa517b3be241ddb892d/1629654183409/fotundersokning-vid-diabetes.pdf [accessed 2026-04-10]
21. Dumas JS, Redish JC. A Practical Guide to Usability Testing: Intellect books; 1999.
22. Adlin T, Pruitt J, Goodwin K, et al. Putting personas to work. Presented at: CHI EA '06: CHI '06 Extended Abstracts on Human Factors in Computing Systems; Apr 22-27, 2006. [doi: 10.1145/1125451.1125456]
23. Reimann R, Cooper A, Cronin D, Noessel C. About Face: The Essentials of Interaction Design: John Wiley & Sons; 2014. URL: https://www.researchgate.net/publication/318707875_About_Face_The_Essentials_of_Interaction_Design_4th_Edition [accessed 2026-04-06]
24. Soegaard M. The basics of user experience design. : Interaction Design Foundation; 2018 URL: https://bpb-eu-w2.wpmucdn.com/sites.aub.edu.lb/dist/c/13/files/2019/06/the-basics-of-ux-design.pdf [accessed 2026-04-06]
25. Lucidchart. 2025. URL: https://www.lucidchart.com/pages [accessed 2025-01-31]
26. Figma. 2025. URL: https://www.figma.com [accessed 2025-02-08]
27. Brooke J. SUS-a quick and dirty usability scale. In: Usability Evaluation in Industry: Taylor & Francis; 1996:189-194 URL: https://www.researchgate.net/publication/319394819_SUS_--_a_quick_and_dirty_usability_scale [accessed 2026-04-06]
28. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]

XSL•FO
RenderX

29. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum –Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
30. Lewis JR, Sauro J. Item benchmarks for the system usability scale. J Usability Stud 2018;13(3):158-169 [FREE Full text]
31. Borsci S, Macredie RD, Martin JL, Young T. How many testers are needed to assure the usability of medical devices? Expert Rev Med Devices 2014 Sep;11(5):513-525. [doi: 10.1586/17434440.2014.940312] [Medline: 25033757]
32. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. JAMA 2025 Jan 7;333(1):71-74. [doi: 10.1001/jama.2024.21972] [Medline: 39425955]
33. Virzi RA, Sorce JF, Herbert LB. A comparison of three usability evaluation methods: heuristic, think-aloud, and performance testing. Proc Hum Factors Ergon Soc Annu Meet 1993 Oct;37(4):309-313. [doi: 10.1177/154193129303700412]
34. Omda picsara. Omda. 2025. URL: https://omda.com/solutions/connected-imaging/omda-picsara [accessed 2025-02-01]
35. Apple. 2025. URL: https://www.apple.com [accessed 2025-02-01]
36. Morse JM, Niehaus L. Mixed Method Design: Principles and Procedures: Routledge; 2016. URL: https://www.routledge.com/rsc/downloads/9781598742978_chapter_2_final.pdf [accessed 2026-04-06]
37. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
38. Bus SA, Sacco ICN, Monteiro-Soares M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3651. [doi: 10.1002/dmrr.3651] [Medline: 37302121]
39. Garces TS, Araújo A, Sousa GJB, et al. Clinical decision support systems for diabetic foot ulcers: a scoping review. Rev Esc Enferm USP 2024;57:e20230218. [doi: 10.1590/1980-220X-REEUSP-2023-0218en] [Medline: 38362842]
40. Zügner R, Jarl G, Sundberg L, Hellstrand Tang U. Experiences of using a digital tool, the D-foot, in the screening of risk factors for diabetic foot ulcers. J Foot Ankle Res 2022 Dec 13;15(1):90. [doi: 10.1186/s13047-022-00594-9] [Medline: 36514099]
41. Kumbhar S, Bhatia M. Advancements and best practices in diabetic foot care: a comprehensive review of global progress. Diabetes Res Clin Pract 2024 Nov;217:111845. [doi: 10.1016/j.diabres.2024.111845] [Medline: 39243866]
42. de Oliveira TCD, de Oliveira AF, de Castro Araújo L, et al. Digital health technologies for diabetic foot ulcers: a systematic review of clinical evidence, access inequities, and public health integration. Int J Environ Res Public Health 2025 Sep 13;22(9):1430. [doi: 10.3390/ijerph22091430] [Medline: 41007574]
43. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
44. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
45. Hidalgo-Ruiz S, Ramírez-Durán MDV, Basilio-Fernández B, et al. Assessment of diabetic foot prevention by nurses. Nurs Rep 2023 Jan 9;13(1):73-84. [doi: 10.3390/nursrep13010008] [Medline: 36648982]
46. Khairat S, Karar A, Khairat S. Facilitators and barriers to ICU providers' usability of electronic health records screens. Stud Health Technol Inform 2025 May 15;327:348-352. [doi: 10.3233/SHTI250342] [Medline: 40380453]
47. Bitomsky L, Pfitzer E, Nißen M, Kowatsch T. Advancing health equity and the role of digital health technologies: a scoping review. BMJ Open 2025 Jun 26;15(6):e099306. [doi: 10.1136/bmjopen-2025-099306] [Medline: 40578855]
48. Brousseau-Foley M, Blanchette V, Houle J, Trudeau F. Developing an interprofessional decision support tool for diabetic foot ulcers management in primary care within the family medicine group model: a Delphi study in Canada. BMC Prim Care 2024 Apr 20;25(1):123. [doi: 10.1186/s12875-024-02387-4] [Medline: 38643081]
49. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Presented at: Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems; Apr 24-29, 1993 URL: https://www.semanticscholar.org/paper/A-mathematical-model-of-the-finding-of-usability-Nielsen-Landauer/245061d94c228abdc9bda2d9a10679e897526465 [accessed 2026-04-06]
50. Nielsen J. Designing Web Usability: The Practice of Simplicity: New riders publishing; 1999.
51. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
52. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
53. Hellstrand Tang U, Smith F, Karilampi UL, Gremyr A. Exploring the role of complexity in health care technology bottom-up innovations: multiple-case study using the nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool. JMIR Hum Factors 2024 Apr 26;11:e50889. [doi: 10.2196/50889] [Medline: 38669076]
54. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng [accessed 2026-04-29]
55. Region Västra Götaland [Article in Swedish]. 2025. URL: https://mediebank.vgregion.se/fotoweb/archives/ [accessed 2026-04-13]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CONSORT: Consolidated Standards of Reporting Trials
DFU: diabetic foot ulcer
EHR: electronic health record
HCI: human-computer interaction
HCP: health care professional
NDR: National Diabetes Register
SUS: System Usability Scale
UX: user experience
VGR: Region Västra Götaland

Edited by I Steenstra; submitted 06.Oct.2025; peer-reviewed by J Kaur, PD Venkatraman, V Blanchette; revised version received 13.Mar.2026; accepted 13.Mar.2026; published 30.Apr.2026.

Please cite as:
Hellstrand Tang U, Ravichandran S, Candefjord S, Sathyapalan DT, Lakshmanan V
User-Centered Development of a Digital Health Service for Diabetic Foot Ulcer Risk Stratification: Usability Study
JMIR Diabetes 2026;11:e83287
URL: https://diabetes.jmir.org/2026/1/e83287
doi:10.2196/83287

© Ulla Hellstrand Tang, Shivani Ravichandran, Stefan Candefjord, Dipu T Sathyapalan, Vivek Lakshmanan. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 30.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Lived Experience of Adolescents With Type 1 Diabetes Mellitus in Bahir Dar, Ethiopia: Phenomenological Study

Yassin Hussin[1], MSc; Desalegn Awle[2], MSc; Hailemariam Mekonnen Workie[2], MSc; Desalegne Amare[2], PhD

[1]School of Nursing and Midwifery, College of Health and Medical Science, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia
[2]College of Medicine and Health Science, Bahir Dar University, Bahir Dar, Ethiopia

Corresponding Author:
Yassin Hussin, MSc
School of Nursing and Midwifery, College of Health and Medical Science, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia

Abstract

Background: Type 1 diabetes mellitus (T1DM) is one of the most common chronic diseases among adolescents and is posing a threat to health and potentially endangering life. It can have a significant impact on the physical, social, and emotional development of adolescents. Understanding the lived experiences of adolescents with T1DM is crucial for improving their health outcomes and future aspirations. However, there is currently limited research, and existing studies on psychosocial and self-care perspectives do not highlight their lived experiences in Ethiopia.

Objective: This study aimed to explore the lived experiences of adolescents living with T1DM in Bahir Dar, Ethiopia.

Methods: A descriptive phenomenological design was used. The study was conducted at Felege Hiwot Comprehensive Specialized Hospital in Bahir Dar from March to May 2023. Ten adolescents aged 10 - 19 years receiving follow-up care at the hospital were selected using purposive sampling. Data were collected through face-to-face, in-depth interviews using a semistructured guide translated into Amharic. All interviews were conducted by the principal investigator, audio-recorded, transcribed verbatim, and translated conceptually. Thematic analysis was conducted using ATLAS.ti (8.4.24; ATLAS.ti Scientific Software Development GmbH) software, guided by the Colaizzi phenomenological method.

Results: The study included 10 participants (7 female and 3 male), aged 13 - 18 years, with the duration of living with T1DM ranging from 6 months to 11 years. It identified 4 key themes in adolescents managing insulin dependence. First, psychoemotional factors were significant, with participants experiencing stress, anxiety, low self-esteem, and negative emotions such as fear and hopelessness, although some also reported positive feelings from strong social support. Second, supportive factors were crucial, as family, friends, teachers, health care providers, and community health insurance offered essential emotional and practical help. Third, significant challenges included insulin unavailability, financial issues, fasting-related nonadherence, fear of hypoglycemia, and sleep disturbances, all impacting their well-being. Finally, adolescents used various coping strategies, such as avoidance, acceptance, prayer, and social engagement, to manage emotional distress and maintain daily functioning.

Conclusions: Adolescents with T1DM in Bahir Dar face multifaceted challenges that affect their emotional well-being and diabetes management. Despite these difficulties, they demonstrate resilience through various coping strategies and benefit from strong support systems. These findings underscore the need for adolescent-centered diabetes care, improved insulin access, and psychosocial support in both clinical and community settings.

(JMIR Diabetes 2026;11:e71179) doi:10.2196/71179

KEYWORDS

adolescents; adolescent health; Bahir Dar; chronic illness; Ethiopia; lived experience; type 1 diabetes mellitus

Introduction

Background

Diabetes mellitus (DM) is a metabolic disorder that causes chronic hyperglycemia due to abnormalities in insulin secretion or action [1]. This condition leads to pathophysiologic changes in multiple organ systems, with 4 main categories, including type 1 DM (T1DM), type 2 DM, gestational DM, and specific types resulting from other causes [2]. T1DM, previously known as insulin-dependent diabetes, is a common chronic childhood disease with peak diagnosis in the midteens [3]. Its exact cause is unknown. T1DM results from autoimmune β-cell destruction, usually leading to absolute insulin deficiency, including latent autoimmune diabetes in adults [4,5].

Adolescence is a transitional phase between childhood and adulthood that involves physical, cognitive, social, and emotional changes leading toward independence [6]. The World

XSL•FO
RenderX

Health Organization (WHO) defines adolescents as individuals aged 10-19 years. [7].

In the past 20 years, the global prevalence of diabetes has nearly tripled, with a significant increase in children and adolescents living with the disease [8]. Currently, there are approximately 50,600 children and adolescents under the age of 20 living with T1DM in Africa [9]. Ethiopia ranks fourth in the number of people with diabetes [8], and the estimated incidence of T1DM among children and adolescents is 2.4 per 100,000 per year [10].

T1DM in adolescents has negative impacts on their daily activities, personal aspirations, and psychological well-being [11]. It can lead to embarrassment, discrimination, limited social relationships, poorer diabetes management, cognitive decline, and a higher risk of developing cardiovascular problems [12]. The unique developmental nature of adolescence, including physical, cognitive, and psychosocial changes, complicates diabetes management for this population. Therefore, achieving ideal metabolic control can be challenging while balancing their developmental needs [13].

Adolescents with T1DM view their condition as a challenge and strive to lead normal lives [14]. Some hope for a cure [15], while others feel different because they have to follow a care plan [16]. Perceptions of T1DM in adolescents tend to be negative [17].

According to a qualitative study, teenagers with T1DM report feeling anxious about blood tests and the long-term implications of their condition. Some of them also expressed discomfort or embarrassment about injecting insulin [18].

Children with T1DM tend to experience more sleep problems, which can impact their behavioral and executive functioning [19,20]. The American Diabetes Association's Standards now recognize the importance of evaluating sleep in clinical settings, as there is emerging evidence showing a relationship between sleep quality and glycemic control [9].

Studies have shown that both family and peer support are important for adolescents with T1DM to effectively manage their condition and adjust to living with a chronic illness. Peer support has been effectively linked to improved compliance and better overall outcomes [21]. In addition, research has shown that adolescents with well-controlled diabetes tend to have stronger family bonds and experience less conflict [22]. Improving diabetes care and support within school settings has been shown to enhance diabetes management and quality of life for this population [23]. However, another study conducted in Ethiopia uncovered that some adolescents faced difficult family dynamics, such as excessive parental control and constant warnings about diabetes management, which created tension [24,25]. Another qualitative study revealed that acceptance, avoidance, and adaptive coping strategies were commonly used by adolescents with T1DM [26].

While there is some research on the lived experiences of individuals with T1DM in Ethiopia, such as studies conducted in North Eastern Ethiopia examining psychosocial and economic perspectives [27] and self-care practices [28], there is still a gap in understanding the specific lived experiences of adolescents with T1DM. This study aimed to fill this gap by exploring the unique experiences of adolescents living with T1DM in this region. The findings are essential for providing a platform for participants to voice their concerns and needs, ultimately leading to improved care and quality of life for both themselves and future adolescents with T1DM. Additionally, it will assist program developers in formulating strategies tailored to adolescents living with T1DM.

Objective

The objective of this study was to explore the lived experiences of adolescents living with T1DM in Bahir Dar, Ethiopia, focusing on their emotional, social, and daily life experiences while managing the disease.

Methods

Study Setting and Period

The study was conducted at Felege Hiwot Comprehensive Specialized Hospital (FHCSH), located in Bahir Dar City, Amhara region, Ethiopia. The hospital was established during the regime of Emperor Haileselassie in April 1963 and provides services for neighboring regions, including Benishangul-Gumuz, serving more than 10 million people.

The department of internal medicine at FHCSH has both inpatient and outpatient units. The hospital also has outpatient follow-up clinics for DM. The chronic follow-up clinic was staffed with internists, residents, general practitioners, and a nurse. Follow-up clients for the diabetes clinic were scheduled for Monday and Tuesday but could also be seen on other days depending on their appointments. On average monthly, 500 - 600 patients with chronic conditions were seen. In addition, 67 children were being followed up for diabetes at the time of data collection. The study was conducted from March to May 2023.

Study Design and Characteristics of Participants

This study used a descriptive phenomenological design to explore the lived experiences of adolescents with T1DM.

Participant Selection and Sampling Strategy

The total number of participants was 10. Participants were selected using purposive sampling, guided by specific inclusion criteria to ensure rich, relevant data. Eligible participants were adolescents aged 10 - 19 years, diagnosed with T1DM for at least 6 months, receiving follow-up care at FHCSH, and able to communicate their experiences. Adolescents with cognitive impairments or acute medical conditions that could interfere with the interview process were excluded. Potential participants were initially identified through the FHCSH diabetes clinic registry, which tracks known patients with diabetes who attend regularly scheduled monthly follow-up appointments. When an identified, eligible adolescent arrived at the hospital for their routine monthly visit, they were contacted and directly approached in person by the principal investigator (PI). The PI introduced the study, explained its purpose, and screened the adolescents against the eligibility criteria. Enrollment was finalized on-site after obtaining written informed consent (and assent from the minor, where applicable) prior to conducting

XSL•FO
RenderX

the interviews. Recruitment continued until data saturation was achieved, defined as the point at which no new themes or insights emerged from subsequent interviews.

Data Collection Method and Procedures

The data were collected through face-to-face in-depth interviews using semistructured, open-ended, guiding questions. The interview guide was developed in English, which comprised sociodemographic, psychological, social, emotional, and economic impact (Multimedia Appendix 1). The question guide translated into Amharic for better understanding. Participants were assigned study codes (P1, P2, P3, etc) based on the order of interviewing. The average interview time ranged from 30 to 60 minutes and included questions about the patient's sociodemographic background and experiences with T1DM as adolescents.

The PI conducted all the interviews. Participants who came for follow-up, communicated with nurses before or after they received treatment in the follow-up room. Meetings were scheduled at times convenient for participants once they agreed to participate in the study. All interviews, except for one, took place at the participant's home.

At the beginning of each interview, the researcher and assistant introduced themselves and explained the study's objectives. Participants were informed that their participation was voluntary and that they could decline or withdraw at any time. The interviews began after participants provided written informed assent (and parental consent, where applicable), completed a brief demographic questionnaire, indicated that they were comfortable proceeding, and granted permission for the interview to be audio-recorded for verbatim transcription. The PI interviewed one adolescent every 3 days, and during the interview, no participant showed signs of distress, and everyone continued to talk in a relaxed manner. The audio recorder was checked to ensure uninterrupted recording.

Upon completion of the questions, the PI asked if there was anything else the participant wanted to discuss. The interviewer then thanked the participant and arranged for follow-up contact to ensure accurate transcription.

Trustworthiness

To ensure the quality of the research findings, the investigators considered different sets of criteria focusing on the credibility, dependability, transferability, and confirmability of the study using various techniques. Prior to data collection, the investigators familiarized themselves with the study setting and established good relationships with relevant staff members. They spent significant time on-site and engaged extensively with the participants. The PI conducted comprehensive field notes, audio recordings, and verbatim transcriptions, which were then reviewed by the coauthor (DA). Additionally, some participants were asked to review the extracted codes to ensure alignment with their experiences and perspectives. The data were described in detail, considering the context of the setting and the purpose of the study. The methods used to procure the data were clearly outlined. Direct quotes from the data were used to validate the findings. A concise summary of the major findings and their implications within the field was provided, with extensive consideration given to limitations and potential areas for future research. The authors sent all collected data tools, raw data, codes generated during the analysis phase, and drawn inferences to researchers not involved in the study for an external audit. The authors securely stored the data. The study ensured that reflexivity and investigators' own opinions did not affect the data and its interpretation. By reviewing transcript data precisely, comparing codes with the raw data, and checking the findings with the participants' views several times, the research process was confirmed by the coauthors.

Researchers' Position

The researchers involved in this study have no personal connection to the phenomenon being studied (T1DM). However, they all come from health care backgrounds and possess a minimum of a master's degree in pediatrics and child health nursing. The authors have a high level of experience in qualitative research, although the extent of their experience varies among researchers.

Data Management and Analysis

Data management and analysis occurred iteratively, commencing immediately following the initial interview to allow preliminary insights to inform subsequent data collection. Audio recordings were transcribed verbatim in Amharic by the PI and translated into English for analysis; back-translation was performed on a subset of transcripts by an independent language expert to ensure semantic equivalence. To guarantee the reproducibility of the study, qualitative data analysis was systematically conducted using ATLAS.ti (version 8.4.24; ATLAS.ti Scientific Software Development GmbH) software, strictly adhering to Colaizzi 7-step descriptive phenomenological framework.

In the first step, the PI repeatedly read the transcripts while listening to the audio files to immerse themselves in the data and gain an intimate understanding of the participants' contexts. Second, the PI reviewed the transcripts line-by-line within ATLAS.ti to isolate phrases and sentences directly pertaining to the adolescents' experiences of living with T1DM. Third, the PI articulated the underlying psychological meanings of each significant statement, abstracting the raw data into conceptual insights without stripping away the participant's original context. Fourth, these formulated meanings were grouped into clusters of themes and subsequent subthemes, which were continuously cross-referenced with the original transcripts to ensure validation. Fifth, the PI integrated the clustered themes into a comprehensive, detailed narrative that fully described the lived experience of adolescent diabetes management. Sixth, this exhaustive description was synthesized and condensed into a succinct statement capturing the essential structure (the "essence") of the phenomenon. Seventh, member-checking was performed by sharing the preliminary summary of findings with a subset of the participants during their subsequent monthly hospital follow-up visits to ensure the final analysis accurately reflected their true lived realities.

Methodologically, the entire analysis was rigorously anchored in Husserlian descriptive phenomenology, which demands capturing the phenomenon as purely experienced by the participants. To achieve this and eliminate researcher bias, the

XSL•FO
RenderX

PI actively practiced Husserlian phenomenological bracketing (epoche) throughout the study. Before and during the analysis, the PI explicitly identified and set aside their own preexisting clinical assumptions, biases, and emotional responses stemming from their professional experience in health care. This bracketing was operationalized by maintaining a detailed reflexive journal, where the PI documented personal reflections after each interview. Furthermore, to enhance trustworthiness, the initial coding and thematic structure generated by the PI were independently reviewed and verified by a coinvestigator experienced in qualitative methods; any coding discrepancies were resolved through collaborative peer debriefing until a 100% consensus on the final thematic framework was reached.

Ethical Considerations

This study was reviewed and approved by the Institutional Review Board (IRB) of Bahir Dar University, College of Medicine and Health Sciences (approval number: BDU/IRB/2023/041). The research involved human subjects and adhered to ethical principles outlined in the Declaration of Helsinki. A permission letter was obtained from FHCSH. Informed assent and consent were obtained from all participants and their legal guardians prior to data collection. Participants were informed about the study's purpose, procedures, potential risks, and their right to withdraw at any time without penalty. Consent included permission to audio-record interviews for transcription and analysis. No secondary data analysis was conducted; all data were collected directly from participants for this study. To ensure privacy and confidentiality, all interview data were anonymized during transcription. Participants were assigned codes (eg, P1 or P2) instead of using personal identifiers. Audio recordings, transcripts, and field notes were securely stored and accessible only to the research team. Data were reported in aggregate form, and direct quotes were deidentified to prevent recognition. No financial compensation was provided to participants. However, interviews were scheduled at times and locations convenient to them, and refreshments were offered when appropriate to ensure comfort and respect for their time.

Results

Sociodemographic Characteristics of the Participants

A total of 10 adolescents living with T1DM participated in this study, including 7 females and 3 males. The participants were assigned codes from P1 to P10. Participants were aged 10-19 years. Seven of them resided in urban areas, while 3 lived in rural areas. Their educational attainment ranged from primary school to preparatory school. The participants had been living with DM for periods ranging from 6 months to 11 years (Table 1).

Table . Background characteristics of the participants of the research of lived experience of adolescents living with type 1 diabetes mellitus (T1DM), Bahir Dar, 2023.

Code	Age (years)	Educational status	Health insurance
P1	17	Grade 11	Yes
P2	15	Grade 7	Yes
P3	16	Grade 9	Yes
P4	18	Grade 12	Yes
P5	16	Grade 9	No
P6	13	Grade 6	Yes
P7	17	Grade 12	Yes
P8	16	Grade 9	Yes
P9	18	Grade 8	No
P10	15	Grade 9	Yes

The Essence of the Lived Experience

The overarching essence of the lived experience for adolescents navigating T1DM within this context is characterized by a fragile, continuous negotiation between an encroaching medical identity and the developmental pursuit of adolescent normality. Rather than experiencing their condition as an isolated medical routine, these participants experienced a profound disruption of self-concept, frequently defining themselves primarily as "patients" burdened by feelings of inferiority, body shame, and existential questioning. This vulnerability is deeply exacerbated by systemic and environmental precarity, including severe sleep disruptions and localized structural threats like the absolute unavailability of insulin. To survive this reality, the adolescent experience is not passive; it is an active, multidimensional struggle for equilibrium. Youth leverage a complex matrix of structural safety nets (eg, Community-Based Health Insurance [CBHI]) and intimate social ecosystems (family, school, and peers) to counterbalance their endurance. Ultimately, their lived experience is defined by a reliance on psychological preservation tactics—alternating between conscious avoidance, spiritual surrender to a higher power, and resilient future-oriented aspirations—to reclaim a sense of agency and humanity from a demanding, lifelong pathology.

Themes

In this study, 4 main themes were identified, with several subthemes underneath each theme. These include psychoemotional factors, supportive factors, challenges to DM management and wellness, and coping mechanisms (Table 2).

XSL•FO
RenderX

Table . Summary of themes, subthemes, and codes of adolescent experiences living with type 1 diabetes mellitus (T1DM) who have a follow-up at Felege Hiwot Hospital, Bahir Dar, Ethiopia, 2023.

Themes	Theme 1: psychoemotional factors	Theme 2: supportive factors	Theme 3: challenge to DM[a] management and well-being	Theme 4: coping toward negative psychoemotional factors
Subthemes	• Psychological factors • Negative • Emotional factors • Positive emotional factors	• Community-Based Health insurance (CBHI) • Hospital care and support • Supportive school environment • Good family and friend support	• Challenge in adherence • A challenge to get insulin • Challenge to sleep	• Avoidance coping, acceptance coping, and adopting coping • Social coping • Spiritual coping
Codes	• Stress • Anxiety • Feeling of inferiority • Low self-esteem • Sense of being different from others • Hate • Sadness • Anger • Heartbreak • Frustration • Awfulness • Love • Happiness • Hopefulness • Hopelessness • Fear • Shame	• Financial support • Friend knowledge of symptoms and management • Empathy • Advice • Help • Respect • Safe place • Family love • Psychological support • Create a feeling of equality • Excellent care • Satisfaction	• Fasting • Sleepy • Difficult to awake • Difficult to initiate sleep • Hypoglycemia • Insulin unavailability • Insulin inaccessibility	• Ignore • Sleeping • Hope • Accepting • Adopt it • Draw picture • Watch movies • Positive self-talk • Seeking help • Prayer

[a]DM: diabetes mellitus.

Theme 1: Psychoemotional Factors

Psychological Factors

Most adolescents living with T1DM revealed psychological factors, including stress, anxiety, feelings of inferiority, low self-esteem, and a sense of being different. Adolescents who struggled with these issues found it difficult to function in their daily lives and experienced a range of problems. All of these factors made it harder for adolescents with type 1 diabetes to engage in normal activities, social life, and school performance. This led to negative impacts on their overall well-being, which could even exacerbate their physical symptoms or contribute to the emergence of other problems such as depression. Addressing psychological factors and promoting self-acceptance can improve the well-being of adolescents with T1DM.

These problems stem from the nature of the disease, the burden of insulin injections, and the lifelong need to manage blood sugar levels. This constant medical regimen contributed to the adolescent feeling inferior. For example,

> *I have come to realize that my care is continuous and lifelong; I always need to follow up with appointments at the hospital and take medication to stay alive. This situation makes me feel disgusted, and when I think about it, I feel bad and inferior.* [17-year-old, female]

Another psychological issue concerned how participants perceived themselves as patients. Many participants described defining themselves primarily as patients, which can lead to feelings of dependence and low self-worth. This perception can affect their mental health and quality of life.

> *…I always feel inferior when I think about having a disease. What I couldn't stop dwelling on, and what caused me pain, was the belief that I am inferior because of my diabetes.* [An 18-year-old, male participant]

Negative Emotional Factors

The study revealed that individuals living with diabetes experience negative emotions, including anger, frustration, embarrassment, fear, hopelessness, heartbreak, and shame. Many participants expressed feelings of sadness and questioned why they had this disease. Some felt self-conscious about administering their insulin injections in front of others and avoided doing so in public. Family members were frustrated and distressed when observing their child's daily insulin injections. The fear of complications such as blindness, hypoglycemia, and microvascular and macrovascular damages were common concerns that led to negative emotions.

> *I am always questioning why I'm living with DM, what I did wrong and why God gave me this disease. This make me feel bad ... Sometimes I'm scared,*

XSL•FO
RenderX

> *especially when I consider potential complications like blindness. living with DM, I may experience life threatening accidents caused by falling down due to hypoglycemia. This always frustrates me.* [A 15-year-old, female adolescent]

Positive Emotional Factors

In the study, some adolescents shared positive emotions such as happiness and optimism about their future, which were not defined by their illnesses but by the quality time spent with family and loved ones. They expressed feeling healthy and content when enjoying moments together. One 17-year-old female participant said, "I'm very happy spending time with my family and friends at recreational places, and I feel just as healthy with them."

Theme 2: Supportive Factors

CBHI

Based on the findings of this study, CBHI proved to be a valuable support system for adolescents with T1DM. Almost all participants expressed gratitude for the financial protection provided by CBHI, which helps to minimize the burden of paying out-of-pocket expenses. They also desired coverage of medical costs incurred outside of the country.

> *I don't pay for medicine, only for transportation, thanks to our family health insurance. Without it, medical expenses would be a struggle for both to me and my family. I believe having health insurance is one of the reasons why I can continue to live here.* [A 17-year-old, female participant]

Similarly, a 16-year-old male adolescent stated, "Health insurance is beneficial, but it would be even more helpful if it could provide financial assistance to Ethiopians in need of medical treatment abroad."

Hospital Care and Support

Most participants expressed that the care and counseling provided by health care providers was of high quality. The patients were satisfied with the care and support provided in the facility. The hospital personnel also corrected any mistakes made while injecting insulin and appreciated the patients when they got it right, which helped build trust and confidence in their self-care management.

A 13-year-old female participant said, "I received exceptional care at the facility, which ultimately prevented me from death. The quality of instruction and the support were excellent."

> *The follow-up care at this hospital is excellent, but I also received good care and follow-up at my previous hospital. The staff in this facility are attentive and inquire about my insulin injection technique, correcting any mistakes and acknowledging correct procedures. They also provide education on proper diabetes management, including warnings against consuming sweet foods.* [A 16-year-old, female participant]

Good Family, Friends, and Community Support

Most participants expressed that they have good support from family members, friends, and the community. The family members of adolescents are supportive in reminding them to take insulin, taking necessary precautions, and actively helping them when they are sick. Similarly, their friends play a significant role in supporting them when they are feeling bad and need academic support. Friends help keep them up-to-date with schoolwork and inform teachers about any problems. For example,

> *…My friends love me like their brothers and sisters, and I am love them too. For example, when I have an appointment at the hospital, they pay close attention in class and update me on what I missed. They inform the teacher if I'm having difficulties and act to support me. I never feel alone because they are always with me. My father reminds me when it's time to take my medicine, just like my mother did. They do everything possible for me when I'm sick. Even my sister is there for me....* [A 17-year-old, female participant]

Furthermore, some adolescents mentioned that they received financial support from their relatives and neighbors, which helped them pay for their treatment and made it possible for them to see this day. A 16-year-old female said, "… What's surprising is I'm here now because my neighbor supported me financially at the beginning; I didn't have enough money at that time."

Supportive School Environments

Most of the participants expressed that the school environment was supportive of adolescents living with T1DM. They stated that they felt safe at school, with their classmates and teachers showing empathy toward their conditions. Moreover, almost all participants mentioned that they did not experience any health-related issues at school, which could be attributed to the supportive environment. For example, a 15-year-old male participant said, "There are no challenges for adolescents living with diabetes to attend school and I feel safe and healthy."

Another 13-year-old female said, "…both my classmates and teachers always support me when I need help" (P6).

Additionally, some adolescents with type 1 diabetes have disclosed their illness to friends at school and feel comfortable talking about it. They have developed close relationships with both male and female classmates and feel positive about their school experience. They have also noted that people at school want to help them, which suggests a supportive culture around chronic illness.

> *I have a good relationship with my classmate, and they are aware of my condition. The school atmosphere is welcoming, and everyone is willing to help me.* [A 16-year-old, male participant]

XSL•FO
RenderX

Theme 3: Challenges to DM Management and Well-Being

Challenges Toward Adherence

Even though all participants reported adhering to their medication most of the time, they faced challenges in doing so. While fasting was not mandatory, some participants attempted it, viewing it as harmless. They fasted without taking their morning injections, which highlighted how fasting could challenge their adherence to insulin.

A 16-year-old female participant shared, "During fasting season, I have never taken my morning medication. I assumed it wouldn't harm me, so I continued fasting."

Some participants mentioned missing doses when traveling without their insulin. An 18-year-old female participant said, "Sometimes I forget my medication when I travel to places like holy water without my insulin…"

Moreover, some participants struggled with adhering to their insulin injections due to the fear of hypoglycemia from demanding activities. This fear stemmed from the severe side effects of hypoglycemia, such as dizziness, weakness, and heart palpitations, which can be alarming. Consequently, they might avoid injecting insulin during extended travel or activities that could trigger hypoglycemia. For example,

> *If I'm traveling for an hour, I skip my insulin because injecting it before or during long journeys could cause a sudden drop in blood sugar levels, leading to dizziness and heart palpitations, which is scary.* [A 17-year-old female participant]

Some participants also mentioned adjusting their medication schedules to accommodate their study routines. An 18-year-old male participant stated, "I sometimes delay my nighttime dose to match my study schedule."

Challenges to Getting Insulin

Most of the participants living with T1DM expressed concerns about the unavailability and inaccessibility of insulin. This had serious consequences for individuals with diabetes, especially during crises such as the COVID-19 pandemic. The lack of access to insulin made it challenging for individuals to effectively manage their diabetes, resulting in various negative health outcomes.

Furthermore, the financial burden of obtaining insulin in areas where it is available can be significant, adding to the challenges faced by individuals with diabetes.

> *The most pressing issue is the availability of insulin, which is a major challenge. The Corona period was very difficult for us because we struggled without insulin.* [An 18-year-old, male adolescent]

> *I had planned to seek medical treatment at X Hospital. However, the hospital staff informed me that they did not currently have insulin available, making it hard for me to receive treatment there. This situation imposes a significant financial burden on me, as I typically spend around 300 birrs on transportation alone, in addition to other expenses like food. If insulin were available at that hospital, it would greatly reduce the financial strain I am currently experiencing.* [A 15-year-old, male participant]

Challenges to Sleeping

Almost all of the adolescent participants living with type 1 diabetes experience sleep problems, such as sleepiness, difficulty waking up in the morning, and trouble initiating sleep. These issues can exacerbate their emotional challenges, affecting their overall well-being. The sleep problems are often caused by the physical symptoms of diabetes, such as frequent urination, thirst, and hunger, as well as the emotional stress of the disease. Lack of sleep can lead to irritability and difficulty concentrating, making it harder to maintain a healthy lifestyle. An 18-year-old female said, "I sleep too much, which makes me feel lethargic. Sometimes, waking up in the morning is difficult for me…"

> *I experience difficulty falling asleep because of unidentifiable concerns that run through my mind. I feel unable to talk about this matter with anyone.* [A 16-year-old, male participant]

Theme 4: Coping Toward Negative Psychoemotional Factors

Adaptation, Acceptance, and Avoidance Coping

Most participants mentioned avoidance, adoption, and acceptance as major coping methods for their psychoemotional problems. Some actively choose to adopt a positive and hopeful mindset through self-talk. For example, a 13-year-old female participant envisioned a future career as a scientist making diabetes medicine, stating, "…I tell myself that I want to be a scientist in the future and develop medicine for diabetes whenever I experience psychoemotional problems."

Others mentioned accepting their circumstances and limitations, recognizing that some things were beyond their control. Participants expressed understanding and a sense of peace regarding their situation, allowing them to move forward without worry. A 16-year-old female participant said, "I understand that it's not my fault. I have given it to the creator, as he knows about his gift, and now I don't have to worry about it."

On the other hand, some participants choose to avoid or ignore their problems by sleeping or engaging in distractions like drawing or watching a movie. A 17-year-old female participant mentioned, "When I feel angry or experience any negative emotions, I cope by drawing a picture and watching a movie."

Similarly, 15-year-old female participants said, "When I experience negative emotions or stress, I tend to ignore them and go to sleep."

Social Coping Mechanism

Nearly half of the participants use social coping strategies as a means of dealing with their psychoemotional problems. These coping mechanisms include seeking support from friends and family, consulting with trusted individuals, and engaging in activities with others. This participant focused on building relationships and finding comfort in the company of others, rather than trying to solve the problem alone. This is supported by:

XSL•FO
RenderX

> *If I am experiencing stress, anxiety, anger, fear or any other negative emotions, I find it helpful to spend time with my friends to relax and feel better. Together, we might take a shower, engage in activities, and ultimately alleviate the negative emotions I am experiencing.* [A 15-year-old, male participant]

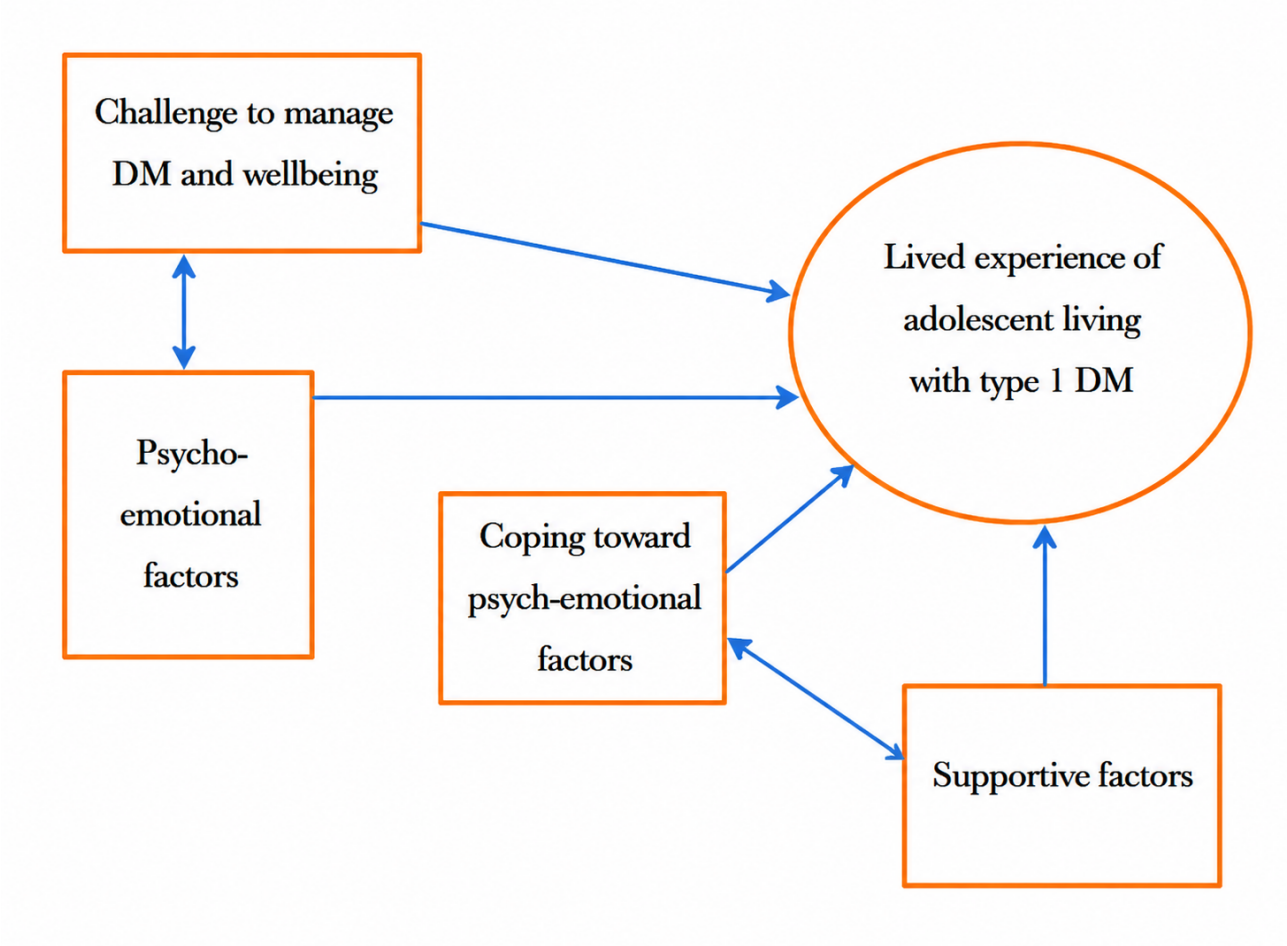
Additionally, an 18-year-old female said, "If I'm experiencing any emotional instability, I find that talking or playing with my friends, makes me feel good."

Spiritual Coping

Some participants explained that prayer helped them relive emotional distress and use it as a coping mechanism. A 16-year-old male participant expressed his experience as, "… if I have any negative emotions, I will go to the church and pray to the God for peace and guidance."

The results showed that the themes and their subthemes had a clear link. As a result, researchers developed a model that shows how each topic in the overall lived experiences of teenagers connects to the others (Figure 1)

Figure 1. A model that shows how each topic in the overall lived experiences of teenagers connects to the others. DM: diabetes mellitus.



Discussion

Principal Findings

Understanding the lived experience of adolescents living with T1DM from their own perspective is a crucial foundation for improving care, particularly in low-resource settings where social, economic, and health-system constraints shape daily disease management. This phenomenological study explored the lived experiences of adolescents with T1DM in Bahir Dar, Ethiopia, a context that is rarely represented in the global diabetes literature. Four interrelated themes emerged, including psychoemotional experiences, supportive factors, challenges to diabetes management and well-being, and coping with negative psychoemotional factors. Together, these themes provide insight into how adolescents navigate T1DM within a setting characterized by limited resources, strong family and community ties, and variable access to specialized diabetes care.

The findings of this study reveal that adolescents living with T1DM experience feelings of inferiority and a sense of being different from their peers, often perceiving themselves as inadequate or less capable in managing their condition. These perceptions may undermine self-esteem and negatively influence both social participation and self-management behaviors. Similar findings have been reported in previous studies, which show that adolescents with type 1 diabetes frequently experience social isolation due to physical limitations, the visibility of treatment routines, and stigma associated with their condition, leading to feelings of exclusion or rejection by peers [29]. In addition, participants described a wide range of emotional experiences, including sadness, frustration, heartbreak, and emotional exhaustion, reflecting the psychological burden of

XSL•FO
RenderX

living with a chronic condition. The demanding and continuous nature of diabetes self-management—such as insulin administration, blood glucose monitoring, and dietary regulation—can be disruptive to daily life and contribute to emotional distress. These challenges may be further intensified during adolescence, a developmental stage marked by hormonal changes, identity formation, and heightened sensitivity to peer acceptance. Consistent with previous studies, adolescents with type 1 diabetes reported difficulties related to treatment demands, negative perceptions of diabetes, school-related challenges, and uncertainty about the future, all of which can contribute to negative emotional and psychological outcomes [18]. These findings underscore the importance of integrating psychosocial support, emotional counseling, and peer-based interventions into routine diabetes care to address both the emotional and social dimensions of living with type 1 diabetes and to support adolescents in developing resilience and effective coping strategies.

Several supportive factors were identified that may contribute to good glycemic control, improved treatment adherence, effective adjustment to chronic conditions, and enhanced quality of life among individuals living with chronic illnesses. One important supportive factor identified in this study is CBHI, which plays a significant role in assisting individuals to manage long-term conditions such as diabetes. This support is particularly crucial for patients who require continuous medical care, regular follow-up, and lifelong treatment. By reducing financial barriers, CBHI facilitates access to essential health care services and medications that might otherwise be unaffordable, thereby improving continuity and quality of care. Improved access to services through CBHI may also promote better adherence to treatment and more consistent disease monitoring, which are essential for optimal health outcomes. These findings are consistent with evidence from Ethiopia, where studies have reported high levels of patient satisfaction with the overall quality of health care services provided through CBHI, highlighting its potential to strengthen chronic disease management and improve patient outcomes [30].

This study shows that adolescents receive adequate support from family, peers, friends, and schools, which is identified as a supportive factor for disease control and quality of life. Adolescents with diabetes encounter numerous challenges, including managing blood sugar levels, addressing the stigma associated with the condition, and coping with the emotional stresses of living with a chronic illness. Therefore, support from family, community, peers, and school can significantly assist in helping adolescents manage their condition and improve quality of life. A study indicates that greater satisfaction with diabetes self-management support from schools, family, peers, and friends is linked to improved metabolic levels and quality of life in adolescents with T1DM [25]. Our finding contradicts that of another study conducted in Ethiopia, which found difficult family dynamics to support the care [25]. Therefore, the presence of a supportive family is important to improve their health conditions and quality of life.

The findings of this study indicate that adolescents living with type 1 diabetes face significant challenges related to the unavailability and inaccessibility of insulin, particularly during crises such as the COVID-19 pandemic, which negatively affected diabetes management and health outcomes. These findings are consistent with global evidence showing that access to insulin remains limited due to high prices, supply-chain markups, weak national procurement and distribution systems, and market dominance by a small number of multinational manufacturers [31,32]. Despite insulin being essential for survival in type 1 diabetes, insufficient policy attention, inadequate health system preparedness, and lack of financial protection continue to hinder equitable access, especially in low- and middle-income settings [33]. The experiences reported by adolescents in this study, therefore, reflect broader global and systemic failures in ensuring reliable access to insulin, underscoring the urgent need for coordinated policy and health system interventions to protect vulnerable populations during both routine care and public health emergencies.

Another important factor influencing glycemic control among adolescents with T1DM is poor treatment adherence, which may result from competing demands such as school workload, travel, or participation in religious fasting. Fasting carries profound spiritual significance for individuals observing religious practices; however, for adolescents with type 1 diabetes, it can pose serious risks to maintaining stable blood glucose levels. While some studies suggest that adolescents with good glycemic control under close medical supervision may fast safely during Ramadan [34], there is limited research on the physiological effects of Christian fasting on anthropometric and biochemical parameters [35]. The findings of this study highlight that fasting can disrupt treatment adherence and glycemic control, potentially endangering the health of adolescents with type 1 diabetes. These results underscore the need for culturally sensitive education and counseling, enabling adolescents to understand the risks of fasting, implement safe management strategies, and maintain consistent adherence to insulin therapy and self-care routines, even during religious observances.

Our study shows that sleeping disturbances and lack of well-being are common problems faced by adolescents living with T1DM. They experience difficulty falling asleep, staying awake, and increased sleeplessness. Other studies have also shown that adolescents with DM often experience sleeping disturbances [19,36]. Fluctuations in blood sugar levels can lead to nocturnal hypoglycemia or hyperglycemia, disrupting sleep [24]. Furthermore, the stress and anxiety related to manage the illness can also contribute to sleep disturbances [34]. This finding suggests that health care providers should consider the impact of sleep on glycemic control.

This study reveals that the participants used 3 main coping mechanisms, including adaptation, acceptance, and avoidance coping, including social and spiritual coping mechanisms. This is consistent with findings from studies conducted in Zambia, indicating that individuals facing chronic illnesses like diabetes may have shared coping experiences [26]. Taken together, the results highlight the importance of addressing the emotional and psychological needs of adolescents with type 1 diabetes through a holistic approach that incorporates personal, social, and spiritual dimensions, alongside medical care, to support resilience, adherence, and overall quality of life.

XSL•FO
RenderX

This study has several limitations that should be considered. First, it was conducted in a single hospital setting, which may limit the transferability of findings to other regions or health care environments. To mitigate this, researchers spent extended time in the field, built rapport with participants, and provided rich contextual descriptions. Future studies could include multiple sites to capture broader perspectives. Second, the sample size was relatively small, comprising 10 participants selected through purposive sampling. Although data saturation was achieved, the limited number may not fully reflect the diversity of adolescent experiences across Ethiopia. Maximum variation sampling was used to include participants of different ages, genders, residences, and duration of illness, but larger, more diverse samples could strengthen generalizability in future research. Third, additional clinical and contextual variables, such as comorbidity, type of insulin, blood glucose monitoring methods, hemoglobin A_{1c} (HbA_{1c}) levels, BMI, and family history, were not collected, though they could provide deeper insight into adolescents' experiences with T1DM. Despite these limitations, the study contributes meaningfully to adolescent-centered diabetes care and highlights the importance of emotional well-being, consistent insulin access, and culturally sensitive management strategies.

Conclusion

This study captures the holistic essence of adolescents with T1DM at Bahir Dar, Ethiopia, exposing a reality where clinical management is inextricably bound to psychological survival and socioeconomic vulnerability. This study makes a unique contribution to the global phenomenology of chronic illness by shifting the narrative away from high-resource Western contexts to showcase how adolescent identity negotiation occurs within the unique cultural, economic, and structural realities of regional Ethiopia. It demonstrates that the financial protection of CBHI and strong collective social networks are vital; yet, they remain vulnerable to systemic failures like medication supply-chain collapses. There is an urgent policy and clinical implication to establish dedicated, multidisciplinary adolescent endocrine clinics within regional referral hospitals. These specialized spaces must formally integrate routine mental health screenings, peer-led support networks, and localized nutritional counseling into the standard monthly follow-up track, transforming diabetes care from an episodic medical intervention into a comprehensive, youth-centered developmental support system.

Acknowledgments

We would like to thank Bahir Dar University College of Medicine and Health Sciences for their support.

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

During the preparation of this manuscript, the authors used ChatGPT/open AI solely to improve the English language, grammar, syntax, and overall readability of the text. After using this tool, the authors reviewed, edited, and verified all content to ensure accuracy, and take full responsibility for the final integrity of the published work. No generative AI was used to create, analyze, or interpret the study data, results, or core scientific findings.

Funding

The authors have also confirmed that no financial funding was received for the study, authorship, and publication of this article.

Data Availability

The data are available from corresponding author and provided upon reasonable request.

Authors' Contributions

Conceptualization: YH
Data curation: DAw
Formal analysis: YH
Funding acquisition: DAw
Investigation: YH
Methodology: YH
Project administration: DA (lead), HM (equal), DAw (supporting)
Resources: YH (lead), DAw (supporting)
Supervision: DA
Validation: HM
Visualization: DA (lead), HM (supporting)
Writing – original draft: YH (lead), DA (supporting)
Writing – review & editing: YH (lead), DA (supporting), HM (supporting)

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The complete semistructured interview guide used in this study.
[DOCX File, 13 KB - diabetes_v11i1e71179_app1.docx]

Checklist 1
SRQR checklist.
[PDF File, 64 KB - diabetes_v11i1e71179_app2.pdf]

References

1. Balaji R, Duraisamy R, Santhosh Kumar MP. Complications of diabetes mellitus: a review. Drug Invention Today 2019;12(1):98-103 [FREE Full text]
2. American Diabetes Association. 2. Classification and diagnosis of diabetes. Diabetes Care 2017 Jan 1;40(Supplement_1):S11-S24. [doi: 10.2337/dc17-S005]
3. Flier JS, Underhill LH, Eisenbarth GS. Type I diabetes mellitus. N Engl J Med 1986 May 22;314(21):1360-1368. [doi: 10.1056/NEJM198605223142106]
4. Association AD. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021. Diabetes Care 2021 Jan 1;44(Supplement_1):S15-S33. [doi: 10.2337/dc21-S002]
5. Roep BO, Thomaidou S, van Tienhoven R, Zaldumbide A. Type 1 diabetes mellitus as a disease of the β-cell (do not blame the immune system?). Nat Rev Endocrinol 2021 Mar;17(3):150-161. [doi: 10.1038/s41574-020-00443-4] [Medline: 33293704]
6. Adams GR GR, Gullotta TP, Montemayor R, editors. Adolescent Identity Formation: Sage Publications; 1992.
7. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health 2018 Mar;2(3):223-228. [doi: 10.1016/S2352-4642(18)30022-1] [Medline: 30169257]
8. Patterson CC, Karuranga S, Salpea P, et al. Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2019 Nov;157:107842. [doi: 10.1016/j.diabres.2019.107842] [Medline: 31518658]
9. I. URL: nternational Diabetes Federation.IDF diabetes atlas eighth edition.International Diabetes Federation; 2017
10. International Diabetes Federation. IDF diabetes atlas seventh edition. : International Diabetes Federation; 2015 URL: https://globaldiabeteseducation.wordpress.com/wp-content/uploads/2017/03/idf_atlas-2015_uk.pdf [accessed 2026-07-20]
11. King KM, King PJ, Nayar R, Wilkes S. Perceptions of adolescent patients of the “Lived Experience” of type 1 diabetes. Diabetes Spectr 2017 Feb;30(1):23-35. [doi: 10.2337/ds15-0041] [Medline: 28270712]
12. Hilliard ME, Wu YP, Rausch J, Dolan LM, Hood KK. Predictors of deteriorations in diabetes management and control in adolescents with type 1 diabetes. J Adolesc Health 2013 Jan;52(1):28-34. [doi: 10.1016/j.jadohealth.2012.05.009] [Medline: 23260831]
13. Straczkowski M, Kowalska I, Dzienis-Straczkowska S, Kinalski M, Górski J, Kinalska I. The effect of exercise training on glucose tolerance and skeletal muscle triacylglycerol content in rats fed with a high-fat diet. Diabetes Metab 2001 Feb;27(1):19-23. [Medline: 11240441]
14. Marshall K, Martin H, Siarkowski Amer K. Exploring perceptions about insulin dependent diabetes mellitus in adolescent patients and peers. Compr Child Adolesc Nurs 2018 Mar;41(1):25-41. [doi: 10.1080/24694193.2017.1316788] [Medline: 29474802]
15. Aguiar GB, Machado MED, Silva LD, Aguiar RD, Christoffel MM. Children with type 1 diabetes mellitus: the experience of disease. Rev Esc Enferm USP 2021;55:e03725. [doi: 10.1590/S1980-220X2020011803725] [Medline: 34190881]
16. Jonker D, Deacon E, van Rensburg E, Segal D. Illness perception of adolescents with well-controlled type 1 diabetes mellitus. Health Psychol Open 2018;5(2):2055102918799968. [doi: 10.1177/2055102918799968] [Medline: 30245842]
17. Lesage S, Deacon E, Van Rensburg E, Segal D. “It kinda sucks”: illness perception of a group of South African adolescents with type 1 diabetes mellitus. Afr J Prim Health Care Fam Med 2021 Feb 17;13(1):e1-e9. [doi: 10.4102/phcfm.v13i1.2782] [Medline: 33764139]
18. Sullivan-Bolyai S, Bova C, Johnson K, et al. Engaging teens and parents in collaborative practice: perspectives on diabetes self-management. Diabetes Educ 2014;40(2):178-190. [doi: 10.1177/0145721713520568] [Medline: 24470041]
19. Caruso NC, Radovanovic B, Kennedy JD, et al. Sleep, executive functioning and behaviour in children and adolescents with type 1 diabetes. Sleep Med 2014 Dec;15(12):1490-1499. [doi: 10.1016/j.sleep.2014.08.011] [Medline: 25441750]
20. Ji X, Wang Y, Saylor J. Sleep and type 1 diabetes mellitus management among children, adolescents, and emerging young adults: a systematic review. J Pediatr Nurs 2021;61:245-253. [doi: 10.1016/j.pedn.2021.06.010] [Medline: 34182231]

XSL•FO
RenderX

21. Butler N, Quigg Z, Bates R, et al. The contributing role of family, school, and peer supportive relationships in protecting the mental wellbeing of children and adolescents. School Ment Health 2022;14(3):776-788. [doi: 10.1007/s12310-022-09502-9] [Medline: 35154501]
22. Anderson BJ, Miller JP, Auslander WF, Santiago JV. Family characteristics of diabetic adolescents: relationship to metabolic control. Diabetes Care 1981;4(6):586-594. [doi: 10.2337/diacare.4.6.586] [Medline: 7347668]
23. Wilt L. The role of school nurse presence in parent and student perceptions of helpfulness, safety, and satisfaction with type 1 diabetes care. J Sch Nurs 2022 Apr;38(2):161-172. [doi: 10.1177/1059840520918310] [Medline: 32292117]
24. Worku K. Exploring the lived experience of adolescents with type 1 diabetes receiving treatment at Tikur Anbessa Specialized Hospital Addis Ababa, Ethiopia: a qualitative study. 2019.
25. Hapunda G. Coping strategies and their association with diabetes specific distress, depression and diabetes self-care among people living with diabetes in Zambia. BMC Endocr Disord 2022 Aug 28;22(1):215. [doi: 10.1186/s12902-022-01131-2] [Medline: 36031626]
26. Ketema Bogale E, Getachew Desalew E, Tesfa H, Endeshaw D, Guadie Tiruneh M, Ketema Bogale S. Lived experience of people with type 1 diabetes in North East Ethiopia; psycho-social and economical perspective, a phenomenological study. Diabetes Metab Syndr Obes 2023;16:3885-3898. [doi: 10.2147/DMSO.S431914] [Medline: 38054036]
27. Bogale EK, Wondiye H, Debela Y, Fentabil Anagaw T, Worku L, Kebede N. Self-care practice, lived experience of type 1 diabetes mellitus patients at Kemisse General Hospital, North Eastern Ethiopia: phenomenological study. SAGE Open Med 2022;10:20503121221126862. [doi: 10.1177/20503121221126862] [Medline: 36172571]
28. Smith LB, Terry A, Bollepalli S, Rechenberg K. School-based management of pediatric type 1 diabetes: recommendations, advances, and gaps in knowledge. Curr Diab Rep 2019 May 24;19(7):1-6. [doi: 10.1007/s11892-019-1158-x] [Medline: 31127416]
29. Tefera BB, Kibret MA, Molla YB, et al. The interaction of healthcare service quality and community-based health insurance in Ethiopia. PLoS One 2021;16(8):e0256132. [doi: 10.1371/journal.pone.0256132] [Medline: 34411148]
30. Beran D, Ewen M, Laing R. Constraints and challenges in access to insulin: a global perspective. Lancet Diabetes Endocrinol 2016 Mar;4(3):275-285. [doi: 10.1016/S2213-8587(15)00521-5] [Medline: 26857998]
31. Beran D, Lazo-Porras M, Mba CM, Mbanya JC. A global perspective on the issue of access to insulin. Diabetologia 2021 May;64(5):954-962. [doi: 10.1007/s00125-020-05375-2] [Medline: 33483763]
32. Beran D, Hirsch IB, Yudkin JS. Why are we failing to address the issue of access to insulin? A national and global perspective. Diabetes Care 2018 Jun;41(6):1125-1131. [doi: 10.2337/dc17-2123] [Medline: 29784696]
33. Willis TA, Gregory AM. Anxiety disorders and sleep in children and adolescents. Sleep Med Clin 2015 Jun;10(2):125-131. [doi: 10.1016/j.jsmc.2015.02.002] [Medline: 26055860]
34. Sow A, Agne F, eacute,ogo A, et al. Effects of lenten fasting on body composition and biochemical parameters. Int J Res Med Sci 2016;4(11):4724-4729. [doi: 10.18203/2320-6012.ijrms20163709]
35. Cho NH, Shaw JE, Karuranga S, et al. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
36. Martyn-Nemeth P, Phillips SA, Mihailescu D, et al. Poor sleep quality is associated with nocturnal glycaemic variability and fear of hypoglycaemia in adults with type 1 diabetes. J Adv Nurs 2018 Oct;74(10):2373-2380. [doi: 10.1111/jan.13765] [Medline: 29917259]

Abbreviations

CBHI: Community-Based Health Insurance
DM: diabetes mellitus
FHCSH: Felege Hiwot Comprehensive Specialized Hospital
HbA_{1c}: hemoglobin A_{1c}
IRB: Institutional Review Board
PI: principal investigator
T1DM: type 1 diabetes mellitus
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Stone, N Cahill; submitted 12.Jan.2025; peer-reviewed by AN Ali, S Birhanu; revised version received 04.Jun.2026; accepted 04.Jun.2026; published 05.Aug.2026.

Please cite as:
Hussin Y, Awle D, Mekonnen Workie H, Amare D
The Lived Experience of Adolescents With Type 1 Diabetes Mellitus in Bahir Dar, Ethiopia: Phenomenological Study
JMIR Diabetes 2026;11:e71179
URL: https://diabetes.jmir.org/2026/1/e71179
doi:10.2196/71179

© Yassin Hussin, Desalegn Awle, Hailemariam Mekonnen Workie, Desalegne Amare. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 5.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Age- and Sex-Specific Distribution of the Triglyceride-Glucose Index in a Large Chinese Population: Cross-Sectional Study

Hanze Du[1*], MD, PhD; Shuaihua Song[1*], MD; Xiao Zhai[1], MD; Junchi Liu[1], MD; Yue Jiang[1], MD; Yaoda Hu[2,3], PhD; Huijing He[2,3], PhD; Daowei Li[4], MD, PhD; Huijuan Zhu[1], MD, PhD; Shi Chen[1], MD, PhD; Guangliang Shan[2,3], PhD; Hui Pan[1], MD, PhD

[1]Key Laboratory of Endocrinology of National Health Commission, Department of Endocrinology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 Shuaifu Garden, Beijing, China

[2]Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & School of Basic Medicine, Peking Union Medical College, Beijing, China

[3]State Key Laboratory of Common Mechanism Research for Major Diseases, Beijing, China

[4]Jilin Provincial Key Laboratory of Tooth Development and Bone Remodeling, School and Hospital of Stomatology, Jilin University, Changchun, Jilin, China

[*]these authors contributed equally

Corresponding Author:
Hui Pan, MD, PhD
Key Laboratory of Endocrinology of National Health Commission, Department of Endocrinology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 Shuaifu Garden, Beijing, China

Abstract

Background: The triglyceride-glucose (TyG) index has demonstrated promising predictive capability in clinical studies, but its distribution characteristics across different age and sex groups in the Chinese population have not been fully characterized.

Objective: This study aimed to describe the population-based distribution of the TyG index.

Methods: A total of 4621 participants aged 20 - 80 years from the China National Health Survey were included in this study. The TyG index was calculated from fasting blood glucose and triglycerides. The age- and sex-specific distribution values of the TyG index were obtained using the percentiles method.

Results: Males had higher BMI (23.32, SD 2.63 vs 22.72, SD 2.60 kg/m^2), triglyceride (133.78, SD 78.07 vs 111.58, SD 61.18 mg/dL), fasting glucose (97.05, SD 11.95 vs 94.34, SD 10.46 mg/dL), and TyG index values (8.63, SD 0.54 vs 8.45, SD 0.50) than females. The TyG index of males reached its peak value at approximately 40 years of age. The lower limit percentile values for females exceeded that of males around age 50. After the age of 60, the upper limit of the distribution values in females was higher than in males.

Conclusions: This study characterized the age- and sex-specific distribution of the TyG index among Chinese adults aged 20 - 80 years. The results of this study contribute to a more precise assessment of glucolipid metabolism.

(JMIR Diabetes 2026;11:e95855) doi:10.2196/95855

KEYWORDS

triglyceride-glucose index; age- and sex-specific distribution; glucolipid metabolism; china national health survey; chinese adults

Introduction

Metabolic syndrome has emerged as a major global health challenge, characterized primarily by insulin resistance, hyperglycemia, dyslipidemia, abdominal obesity, and hypertension [1]. These components are not only significant risk factors for chronic conditions such as cardiovascular disease, but also contribute to increased mortality rates [2]. Early detection and intervention for insulin resistance, a core feature of metabolic syndrome, are crucial to slowing disease progression [3]. However, traditional methods for assessing insulin resistance, such as the Homeostasis Model Assessment of Insulin Resistance, often require complex laboratory testing and incur high costs [4]. Consequently, the exploration of simple and cost-effective indicators to evaluate insulin resistance in clinical practice is of great importance.

In recent years, the triglyceride-glucose index (TyG index) has gained attention for its ability to effectively estimate insulin resistance. The TyG index is a simple, economical screening tool for this purpose [5-7]. Studies have shown that the TyG

XSL•FO
RenderX

index is closely associated with metabolic syndrome, diabetes, cardiovascular disease, and mortality and has the potential to become a next-generation predictor for these conditions [4,8,9].

Although the TyG index has demonstrated promising predictive capability in clinical studies, its distribution characteristics across different age and sex groups have not been fully characterized in the Chinese population. Consequently, understanding population-based distribution patterns of the TyG index may improve interpretation of this marker in different demographic groups. This study aims to characterize the age- and sex-specific distribution of the TyG index in Han Chinese adults, thereby providing population-based distribution data for future research and clinical assessments of glucolipid metabolism.

Methods

Data Source and Study Population

Data for this study were obtained from the China National Health Survey database and included participants who met the criteria in Guangdong, Jilin, and Jiangsu provinces [10]. Exclusion criteria included (1) age <20 years or >80 years; (2) non-Han ethnicity; (3) BMI <14 kg/m^2 or >28 kg/m^2 [11]; (4) missing data for fasting glucose (GLU) or triglyceride (TG); and (5) participants with diabetes mellitus, hypertension, hyperlipidemia, hyperuricemia, other cardiovascular and cerebrovascular diseases, hepatic and renal diseases, thyroid disorders, pancreatic disorders, tumors, and other acute diseases. Information on disease history was assessed based on self-reports using standardized questionnaires. A total of 4621 participants were included in this study.

Data Collection

All participants were required to fast for at least 8 hours, after which fasting blood samples were collected the following morning by professional nurses. The separated plasma or serum was aliquoted before the initial thawing and stored at −80°C to ensure sample stability. Serum TG and GLU levels were measured using a chemical analyzer (Cholesterol Cobas8000C701), and TyG index values were calculated using the formula TyG index=ln(TG [mg/dL] × GLU [mg/dL] / 2) [12]. Additionally, demographic information, including birth date, gender, and history of diabetes, hypertension, hyperlipidemia, hyperuricemia, and other relevant medical conditions, was collected through standardized questionnaires, ensuring the comprehensiveness and accuracy of the data.

Ethical Considerations

This study was conducted according to the principles of the Declaration of Helsinki and approved by the institutional review board of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (number 028‐2013). All participants provided their consent for participation. Participants were informed that their personal information and research data would be kept strictly confidential, securely managed, and used only for research purposes in accordance with the approved study protocol and relevant ethical requirements.

Statistical Analysis

Continuous variables were presented as mean (SD). One-way ANOVA was performed for intergroup comparisons, followed by post hoc multiple comparisons using the Bonferroni method. The age- and sex-specific percentiles (2.5th, 50th, and 97.5th) of the TyG index were calculated directly from the empirical distribution (ie, the ordered values) within each subgroup. The 95% CIs for these percentiles were estimated using the binomial exact method, a robust nonparametric approach. To visualize the age- and sex-specific distribution ranges for the TyG index, we constructed grouped percentile curves. Specifically, after stratifying the sample by sex and 10-year age intervals, we calculated the 2.5th percentile (lower limit) and the 97.5th percentile (upper limit) of the TyG index within each group. The curves were generated by smoothing splines through the midpoints of successive age groups to depict the trend. The scatter points illustrate the distribution of the individual raw data. All statistical analyses were conducted using Stata (version 16.0) and R (version 4.3.1) software, with the “tidyverse” and “ggplot2” packages used for data manipulation and visualization. A 2-sided significance determined as $P<.05$.

Results

A total of 4621 subjects were included in this study, with a total of 1418 males and 3203 females. More baseline characteristics are shown in Table S1 in Multimedia Appendix 1. Figure 1 shows the age- and sex-specific distribution (2.5th to 97.5th percentiles) of the TyG index values. Table 1 and Figure 2 illustrate the age- and sex-specific percentile values and distribution of TyG index values. In this study, the TyG index values of females exhibited a gradual increase with age. In contrast, in males, the TyG index values peaked around 40 years of age and subsequently declined. The lower percentile value of the TyG index for females surpassed that for males at approximately 50 years of age. Furthermore, from 20 to 59 years, females showed lower upper percentile values of the TyG index compared with males. After 60 years of age, the upper percentile values of the TyG index were higher in females than in males. We also analyzed age- and sex-specific TyG index values across the three provinces, with detailed results provided in Table S2 in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. The distribution (2.5th-97.5th percentile) of TyG index values by age and sex. TyG: triglyceride-glucose.

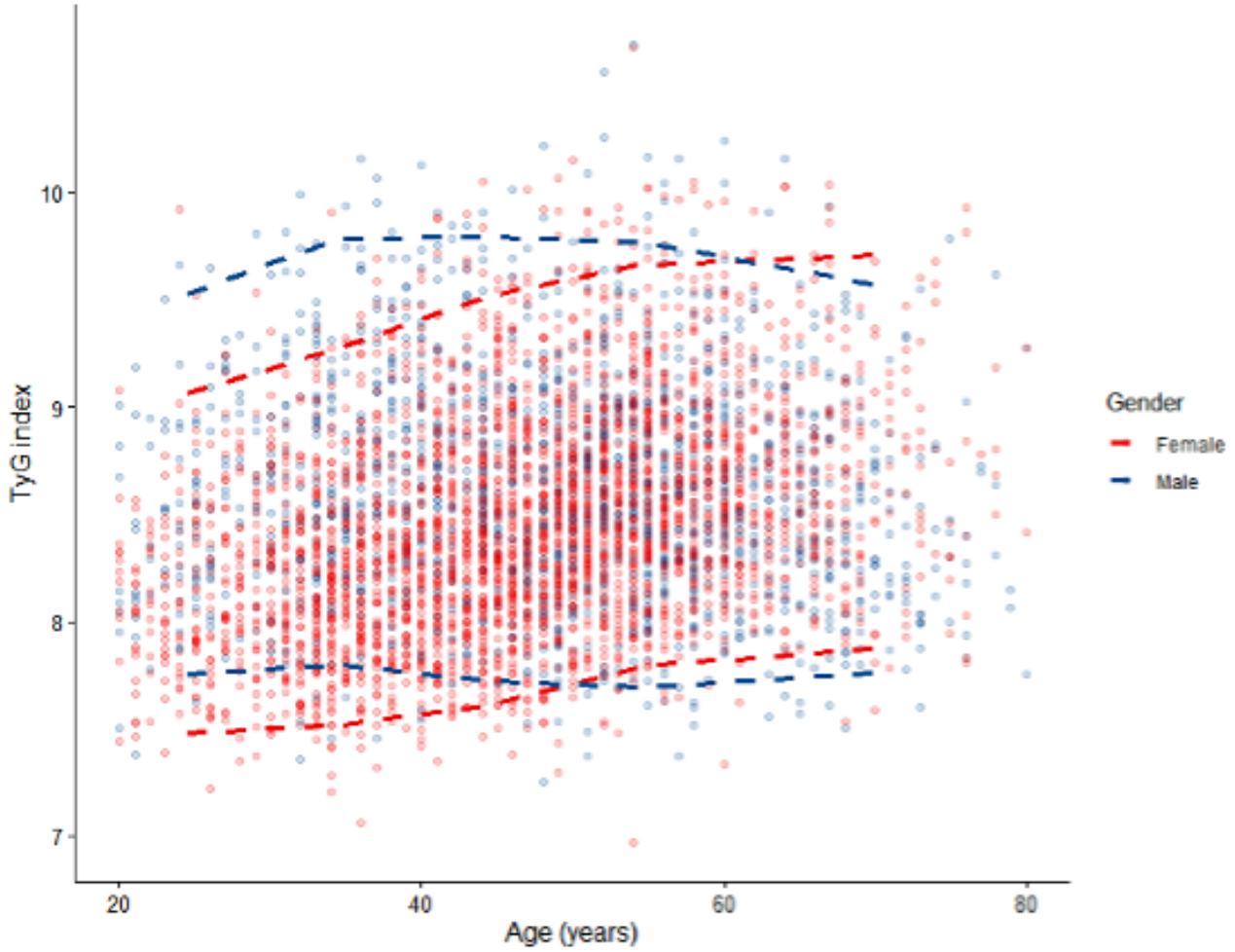
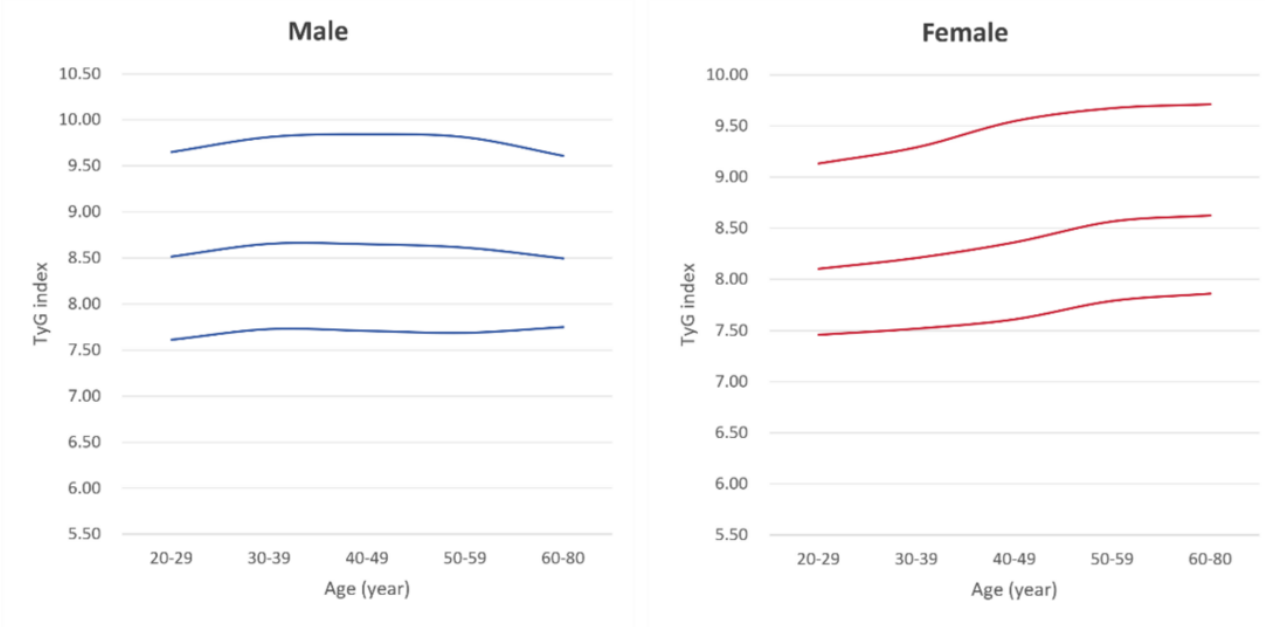


Table . Distribution of triglyceride-glucose index values by age and sex.

	Number of participants	2.5% (95% CIs)	50.0% (95% CIs)	97.5% (95% CIs)
Males, age groups (years)				
20 - 29	115	7.61 (7.38 - 7.88)	8.52 (8.35 - 8.63)	9.65 (9.28 - 9.81)
30 - 39	261	7.73 (7.45 - 7.86)	8.65 (8.55 - 8.74)	9.82 (9.70 - 10.03)
40 - 49	316	7.71 (7.53 - 7.80)	8.65 (8.56 - 8.72)	9.84 (9.69 - 9.98)
50 - 59	403	7.69 (7.59 - 7.73)	8.61 (8.54 - 8.68)	9.81 (9.64 - 10.16)
60 - 80	323	7.75 (7.59 - 7.82)	8.50 (8.42 - 8.57)	9.61 (9.49 - 9.99)
Females, age groups (years)				
20 - 29	251	7.46 (7.36 - 7.54)	8.10 (8.04 - 8.19)	9.13 (8.90 - 9.53)
30 - 39	665	7.52 (7.47 - 7.57)	8.21 (8.15 - 8.25)	9.29 (9.17 - 9.35)
40 - 49	951	7.61 (7.57 - 7.64)	8.36 (8.32 - 8.40)	9.55 (9.41 - 9.63)
50 - 59	912	7.79 (7.73 - 7.83)	8.57 (8.54 - 8.61)	9.68 (9.61 - 9.84)
60 - 80	424	7.86 (7.79 - 7.93)	8.63 (8.55 - 8.70)	9.71 (9.61 - 9.93)

XSL•FO
RenderX

Figure 2. The age- and sex-specific distribution of TyG index values. TyG: triglyceride-glucose.



Discussion

Principal Findings

This study established age- and sex-specific percentile distributions of the TyG index based on 4621 Han Chinese adults. To our best knowledge, this is the first study to characterize age- and sex-specific distribution patterns of TyG index values in a large Chinese population. These findings provide population-based distribution data that may support the interpretation of TyG index levels in future epidemiological and clinical research.

The TyG index has been widely used in the diagnosis and management of various diseases, including cardiovascular and cerebrovascular diseases, metabolic diseases, and liver and kidney diseases [4]. However, the population-based distribution characteristics of the TyG index in apparently healthy individuals remain incompletely characterized. This study found that males had higher BMI, TG, GLU, and TyG index values compared to females, which was consistent with previous research [13-15]. The male participants in this study were slightly older than the females. Given that metabolic indicators such as TG and GLU are positively correlated with age, the higher TyG index values observed in males may be partly attributable to this age difference [16]. Moreover, studies have indicated potential gender differences in metabolic processes, particularly in glucolipid metabolism. These differences may be associated with the modulatory effects of sex hormones on lipid and glucose homeostasis [17].

Subsequently, we analyzed the TyG index values of participants from three different provinces in China by age and gender, and the results showed that healthy males aged 60 - 80 years from Jilin had higher TyG index values than the others. These differences are likely driven by a combination of geographic, economic, lifestyle, and dietary factors. First, at the geographic and climatic level, the three provinces are listed in a north-to-south geographic order: northeastern China (Jilin), eastern China (Jiangsu), and southern China (Guangdong), respectively, with significant differences in latitude, mean annual temperature, and humidity. These variations may indirectly influence glucolipid metabolism by affecting basal metabolic rates, seasonal food availability, and physical activity patterns [18]. Second, socioeconomic and urbanization factors differ across the three provinces in terms of economic development, urbanization rates, and health care accessibility, which may act as potential confounders. Third, dietary factors and environmental conditions may also play a role. Jilin Province, located in the northeastern part of China, experiences long and harsh winters, and the local dietary pattern may be characterized by relatively high energy intake. These factors could potentially contribute to differences in metabolic profiles across regions. Particularly in the older adult population, although participants in this region do not report metabolic-related diseases, age-related decline in metabolic function may be associated with gradual increases in blood lipid and glucose levels [19]. Therefore, the observed differences should be interpreted as the result of synergistic effects of multiple environmental and lifestyle factors. However, these explanations remain speculative, as dietary intake and environmental exposure were not directly measured in the present study. Future studies should collect more comprehensive individual-level data on socioenvironmental and behavioral factors to more accurately elucidate the causes of cardiometabolic risk differences across various regions of China.

Our findings indicated that the median TyG index values peaked in males aged 40 to 49 years. In males older than 50 years, blood glucose levels were slightly elevated and triglyceride levels were significantly lower; consequently, the TyG index values were gradually surpassed by those of females. Although men generally have a higher cardiometabolic risk than premenopausal women, the lower TyG index values observed among males over 50 years in this study may be partly associated with healthier lifestyle and dietary habits among these participants [20,21]. Our results indicate that among individuals aged 50

XSL•FO
RenderX

and above, compared to females, males may need to maintain their TyG index values within a stricter range to sustain a relatively healthy metabolic state. In females, TyG index values increased with age. Furthermore, sex differences in the TyG index distribution varied with age. After approximately 50 years of age, females showed higher lower percentile values than males, and by around 60 years, females also exhibited higher upper percentile values. Current studies have demonstrated that postmenopausal women are at an increased risk of insulin resistance and glucose intolerance, accompanied by a gradual elevation in lipid levels [22,23]. Estrogen plays a crucial regulatory role in glucolipid metabolism, and the abrupt decline in estrogen levels following menopause significantly elevates the risk of metabolic syndrome and cardiovascular diseases [24,25]. Nevertheless, females were generally able to maintain a relatively favorable metabolic condition even at higher TyG index values.

Recent studies have increasingly focused on the association between TyG index thresholds and disease-specific morbidity or mortality risks. A study by Zhang et al [26], which used the National Health and Nutrition Examination Survey database, demonstrated that baseline TyG index values exceeding specific thresholds (TyG index >9.05 for all-cause mortality and >8.84 for cardiovascular mortality) exhibited positive correlations with mortality risk. Similarly, Liu et al [27] identified TyG index inflection points at 9.104 for all-cause mortality and 8.758 for cardiovascular mortality. Zhou et al [28] further found TyG index thresholds (male >8.81, female >8.73) as indicators of insulin resistance in Chinese populations. Notably, the upper percentile values of the TyG index observed in the study population were higher than previously reported cutoff values. This discrepancy may be partly attributed to differences in population characteristics and study design, including variations in inclusion and exclusion criteria across studies. Importantly, previously reported cutoff values were primarily derived from prognostic or risk prediction models for cardiometabolic outcomes, where thresholds are optimized to enhance predictive performance rather than to reflect population-based distribution characteristics. In addition, given the nature of large-scale population-based surveys, our study population may inevitably include individuals with undiagnosed or subclinical metabolic abnormalities. Moreover, to better reflect real-world population structure, individuals with overweight were included, which may also contribute to the observed differences in distribution compared with previously reported thresholds. These findings highlight the importance of characterizing population-specific distribution patterns of the TyG index to improve the interpretation of physiological variation across different demographic groups.

This study has several strengths. First, the inclusion of 4621 participants ensured a sample size that was moderate and statistically significant. Additionally, participants recruited from geographically diverse and representative provinces across China enhanced the reliability of the findings. Limitations should also be acknowledged. The assessment of disease history in this study was primarily based on participant self-reports. While this method serves as a common operational definition in large-scale epidemiological investigations, it may not fully exclude individuals with undiagnosed or subclinical conditions. Due to the limited sample size in the age 70 - 80 subgroup, individuals aged 60 - 80 were combined into a single stratum for analysis. While this ensures more statistically stable percentile estimates, it may obscure important physiological trends within this broad 20-year age span, given the considerable heterogeneity in metabolic parameters among older adults. The precision of the estimated percentile values may be limited in some age-sex subgroups with smaller sample sizes. Furthermore, the distribution characteristics in this study are derived from Han Chinese populations in three provinces—Jilin, Jiangsu, and Guangdong—listed in north-to-south geographic order and representing northeastern, eastern, and southern China, respectively. While the results reflect the characteristics of the Han Chinese in these regions, caution should be exercised when extrapolating them to other geographical areas or ethnic groups across the nation. Further studies are needed to characterize the age- and sex-specific distribution of the TyG index in other ethnic groups and in younger populations, including adolescents.

Conclusion

This study characterized the age- and sex-specific distribution of TyG index values in Chinese adults aged 20 - 80 years using a large national survey. These population-based distribution data provide a reference framework for future epidemiological and clinical research on glucolipid metabolism.

Acknowledgments

We thank all the China National Health Survey participants and all the staff who gave their considerable time and energy to this survey. The artificial intelligence tool Yuanbao was used for language editing of the manuscript to improve language clarity.

Funding

This study was supported by Research on the Basic Resources of Science and Technology in the Ministry of Science and Technology (2022FY100800).

Data Availability

The data are available from the corresponding author on reasonable request.

Authors' Contributions

Writing – original draft: SS.

XSL•FO
RenderX

Methodology: HD, SS.
Formal analysis: HD, SS.
Software: SS.
Investigation and data curation: SS, XZ, JL, YJ, YH, HH.
Writing – review & editing: DL, HD.
Supervision: HZ, SC, GS, HP.
Conceptualization: HD, SS, SC, GS, HP.
Investigation: SS, SC, GS, HP.
Project administration: GS, HP.
Funding acquisition: HD, GS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the study population.
[DOCX File, 16 KB - diabetes_v11i1e95855_app1.docx]

Multimedia Appendix 2
Distribution of triglyceride-glucose index by age and sex in three provinces.
[DOCX File, 19 KB - diabetes_v11i1e95855_app2.docx]

References

1. Silveira Rossi JL, Barbalho SM, Reverete de Araujo R, Bechara MD, Sloan KP, Sloan LA. Metabolic syndrome and cardiovascular diseases: going beyond traditional risk factors. Diabetes Metab Res Rev 2022 Mar;38(3):e3502. [doi: 10.1002/dmrr.3502] [Medline: 34614543]
2. Ciumărnean L, Milaciu MV, Negrean V, et al. Cardiovascular risk factors and physical activity for the prevention of cardiovascular diseases in the elderly. Int J Environ Res Public Health 2021 Dec 25;19(1):207. [doi: 10.3390/ijerph19010207] [Medline: 35010467]
3. Fahed G, Aoun L, Bou Zerdan M, et al. Metabolic syndrome: updates on pathophysiology and management in 2021. Int J Mol Sci 2022 Jan 12;23(2):786. [doi: 10.3390/ijms23020786] [Medline: 35054972]
4. Sun Y, Ji H, Sun W, An X, Lian F. Triglyceride glucose (TyG) index: a promising biomarker for diagnosis and treatment of different diseases. Eur J Intern Med 2025 Jan;131:3-14. [doi: 10.1016/j.ejim.2024.08.026] [Medline: 39510865]
5. Huo RR, Liao Q, Zhai L, You XM, Zuo YL. Interacting and joint effects of triglyceride-glucose index (TyG) and body mass index on stroke risk and the mediating role of TyG in middle-aged and older Chinese adults: a nationwide prospective cohort study. Cardiovasc Diabetol 2024 Jan 13;23(1):30. [doi: 10.1186/s12933-024-02122-4] [Medline: 38218819]
6. Dang K, Wang X, Hu J, et al. The association between triglyceride-glucose index and its combination with obesity indicators and cardiovascular disease: NHANES 2003-2018. Cardiovasc Diabetol 2024 Jan 6;23(1):8. [doi: 10.1186/s12933-023-02115-9] [Medline: 38184598]
7. Son DH, Lee HS, Lee YJ, Lee JH, Han JH. Comparison of triglyceride-glucose index and HOMA-IR for predicting prevalence and incidence of metabolic syndrome. Nutr Metab Cardiovasc Dis 2022 Mar;32(3):596-604. [doi: 10.1016/j.numecd.2021.11.017] [Medline: 35090800]
8. Liu C, Liang D, Xiao K, Xie L. Association between the triglyceride-glucose index and all-cause and CVD mortality in the young population with diabetes. Cardiovasc Diabetol 2024 May 16;23(1):171. [doi: 10.1186/s12933-024-02269-0] [Medline: 38755682]
9. Alizargar J, Bai CH, Hsieh NC, Wu SFV. Use of the triglyceride-glucose index (TyG) in cardiovascular disease patients. Cardiovasc Diabetol 2020 Jan 15;19(1):8. [doi: 10.1186/s12933-019-0982-2] [Medline: 31941513]
10. He H, Pan L, Pa L, et al. Data resource profile: the China National Health Survey (CNHS). Int J Epidemiol 2018 Dec 1;47(6):1734-1735f. [doi: 10.1093/ije/dyy151] [Medline: 30124853]
11. Chen K, Shen Z, Gu W, et al. Prevalence of obesity and associated complications in China: a cross-sectional, real-world study in 15.8 million adults. Diabetes Obes Metab 2023 Nov;25(11):3390-3399. [doi: 10.1111/dom.15238] [Medline: 37589256]
12. Guerrero-Romero F, Simental-Mendía LE, González-Ortiz M, et al. The product of triglycerides and glucose, a simple measure of insulin sensitivity. Comparison with the euglycemic-hyperinsulinemic clamp. J Clin Endocrinol Metab 2010 Jul;95(7):3347-3351. [doi: 10.1210/jc.2010-0288] [Medline: 20484475]
13. Hayes A, Gearon E, Backholer K, Bauman A, Peeters A. Age-specific changes in BMI and BMI distribution among Australian adults using cross-sectional surveys from 1980 to 2008. Int J Obes (Lond) 2015 Aug;39(8):1209-1216. [doi: 10.1038/ijo.2015.50] [Medline: 25869604]

XSL•FO
RenderX

14. Zhang X, Meng Z, Li X, et al. The association between total bilirubin and serum triglyceride in both sexes in Chinese. Lipids Health Dis 2018 Sep 12;17(1):217. [doi: 10.1186/s12944-018-0857-7] [Medline: 30208888]
15. Mauvais-Jarvis F. Gender differences in glucose homeostasis and diabetes. Physiol Behav 2018 Apr 1;187:20-23. [doi: 10.1016/j.physbeh.2017.08.016] [Medline: 28843891]
16. Li J, Liu M, Liu F, et al. Age and genetic risk score and rates of blood lipid changes in China. JAMA Netw Open 2023 Mar 1;6(3):e235565. [doi: 10.1001/jamanetworkopen.2023.5565] [Medline: 36988954]
17. Shi Y, Ma J, Li S, et al. Sex difference in human diseases: mechanistic insights and clinical implications. Signal Transduct Target Ther 2024 Sep 10;9(1):238. [doi: 10.1038/s41392-024-01929-7] [Medline: 39256355]
18. Zhang Y, Tong M, Wang B, et al. Geographic, gender, and seasonal variation of diabetes: a nationwide study with 1.4 million participants. J Clin Endocrinol Metab 2021 Nov 19;106(12):e4981-e4992. [doi: 10.1210/clinem/dgab543] [Medline: 34314489]
19. Fountain WA, Bopp TS, Bene M, Walston JD. Metabolic dysfunction and the development of physical frailty: an aging war of attrition. Geroscience 2024 Aug;46(4):3711-3721. [doi: 10.1007/s11357-024-01101-7] [Medline: 38400874]
20. Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. Nat Med 2019 Nov;25(11):1657-1666. [doi: 10.1038/s41591-019-0643-8] [Medline: 31700185]
21. Regitz-Zagrosek V, Lehmkuhl E, Mahmoodzadeh S. Gender aspects of the role of the metabolic syndrome as a risk factor for cardiovascular disease. Gend Med 2007;4 Suppl B:S162-S177. [doi: 10.1016/s1550-8579(07)80056-8] [Medline: 18156101]
22. Kim SY, Go TH, Lee JH, et al. Differential association of metabolic syndrome and low-density lipoprotein cholesterol with incident cardiovascular disease according to sex among Koreans: a national population-based study. Eur J Prev Cardiol 2022 Feb 9;28(18):2021-2029. [doi: 10.1093/eurjpc/zwaa114] [Medline: 33823542]
23. Godsland IF. Oestrogens and insulin secretion. Diabetologia 2005 Nov;48(11):2213-2220. [doi: 10.1007/s00125-005-1930-0] [Medline: 16193292]
24. Ko SH, Kim HS. Menopause-associated lipid metabolic disorders and foods beneficial for postmenopausal women. Nutrients 2020 Jan 13;12(1):202. [doi: 10.3390/nu12010202] [Medline: 31941004]
25. Mauvais-Jarvis F, Clegg DJ, Hevener AL. The role of estrogens in control of energy balance and glucose homeostasis. Endocr Rev 2013 Jun;34(3):309-338. [doi: 10.1210/er.2012-1055] [Medline: 23460719]
26. Zhang Q, Xiao S, Jiao X, Shen Y. The triglyceride-glucose index is a predictor for cardiovascular and all-cause mortality in CVD patients with diabetes or pre-diabetes: evidence from NHANES 2001-2018. Cardiovasc Diabetol 2023 Oct 17;22(1):279. [doi: 10.1186/s12933-023-02030-z] [Medline: 37848879]
27. Liu Q, Zhang Y, Chen S, et al. Association of the triglyceride-glucose index with all-cause and cardiovascular mortality in patients with cardiometabolic syndrome: a national cohort study. Cardiovasc Diabetol 2024 Feb 24;23(1):80. [doi: 10.1186/s12933-024-02152-y] [Medline: 38402393]
28. Zhou H, Ding X, Lan Y, et al. Dual-trajectory of TyG levels and lifestyle scores and their associations with ischemic stroke in a non-diabetic population: a cohort study. Cardiovasc Diabetol 2024 Jun 28;23(1):225. [doi: 10.1186/s12933-024-02313-z] [Medline: 38943172]

Abbreviations

GLU: fasting glucose
TG: triglyceride
TyG: triglyceride-glucose

Edited by I Steenstra, S Li; submitted 22.Mar.2026; peer-reviewed by H Wu, Q Fan; revised version received 07.May.2026; accepted 10.May.2026; published 27.May.2026.

Please cite as:
Du H, Song S, Zhai X, Liu J, Jiang Y, Hu Y, He H, Li D, Zhu H, Chen S, Shan G, Pan H
Age- and Sex-Specific Distribution of the Triglyceride-Glucose Index in a Large Chinese Population: Cross-Sectional Study
JMIR Diabetes 2026;11:e95855
URL: https://diabetes.jmir.org/2026/1/e95855
doi:10.2196/95855

© Hanze Du, Shuaihua Song, Xiao Zhai, Junchi Liu, Yue Jiang, Yaoda Hu, Huijing He, Daowei Li, Huijuan Zhu, Shi Chen, Guangliang Shan, Hui Pan. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 27.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR

XSL•FO
RenderX

Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Harnessing Digital Innovation for Diabetes Care: Insights From the Action4Diabetes–CorrelAid Data4Good Collaboration

Sze May Ng[1,2], PhD; Michael Aydinbas[3,4], MSc; Tyla Martin[5], BSc; Darren Jia Chen Toh[6]; Steven James[7,8,9], PhD

[1]Faculty of Health Social Care and Medicine, Edge Hill University, St Helens Road, Ormskirk, United Kingdom
[2]Department of Paediatrics, Mersey and West Lancashire Teaching Hospitals, Ormskirk, United Kingdom
[3]CorrelAid e.V., Berlin, Germany
[4]Xibix Solutions GmbH, Unterschleißheim, Germany
[5]Action4Diabetes, Bangkok, Thailand
[6]Cambridge University, Cambridge, Cambridgeshire, United Kingdom
[7]University of the Sunshine Coast, Petrie, Queensland, Australia
[8]University of Melbourne, Parkville, Victoria, Australia
[9]Western Sydney University, Campbelltown, Victoria, Australia

Corresponding Author:
Sze May Ng, PhD
Faculty of Health Social Care and Medicine, Edge Hill University, St Helens Road, Ormskirk, United Kingdom

Abstract

Recent decades have seen a dramatic proliferation of real-world data use and evidence generation from nonresearch settings. Data utilization is particularly revolutionizing the operations and impact of nongovernmental organizations worldwide, especially in low- and middle-income countries. Action4Diabetes, which has incrementally been providing sustainable diabetes care for children, adolescents, and young adults with type 1 diabetes (aged 0 - 25 y) across Southeast Asia since 2015, is one such organization. Recognizing the importance of data, Action4Diabetes have collaborated with CorrelAid e.V. As part of this, Action4Diabetes has been exchanging patient data with the local program hospitals monthly. A preprocessing pipeline was implemented, extracting patient and medical product data in a standardized and unified manner. Data collected are anonymized and subsequently uploaded to secure public cloud storage, where they are processed and stored in a centralized database. The model used by Action4Diabetes shows that much can be achieved and can perhaps be utilized elsewhere.

(JMIR Diabetes 2026;11:e89357) doi:10.2196/89357

KEYWORDS

Action4Diabetes; diabetes; digital innovation; evidence; Southeast Asia

Introduction

Recent decades have seen a dramatic proliferation of the use of real-world data and evidence generation from nonresearch settings, including both health and nonhealth sources [1]. Data utilization is particularly revolutionizing the operations and impact of not-for-profit, nongovernmental organizations worldwide, especially in low- and middle-income countries where national health systems are limited [2]. This effort is vital for understanding health trends; identifying care and best practice gaps; supporting the design and evaluation of community-based sustainable, locally owned health solutions; and informing strategic planning and policy. Recognizing the importance of data, since February 2023, the nongovernmental organization Action4Diabetes [3-6] (Multimedia Appendix 1), which provides sustainable diabetes care to youth across Southeast Asia, has collaborated with CorrelAid e.V., a not-for-profit nongovernmental organization of data scientists [7]. This viewpoint aims to describe the preprocessing data pipeline that was implemented in collaboration, across Southeast Asia, to account for differences between files from different hospitals and over time for the same hospital; the pipeline was written in R and built on GitHub.

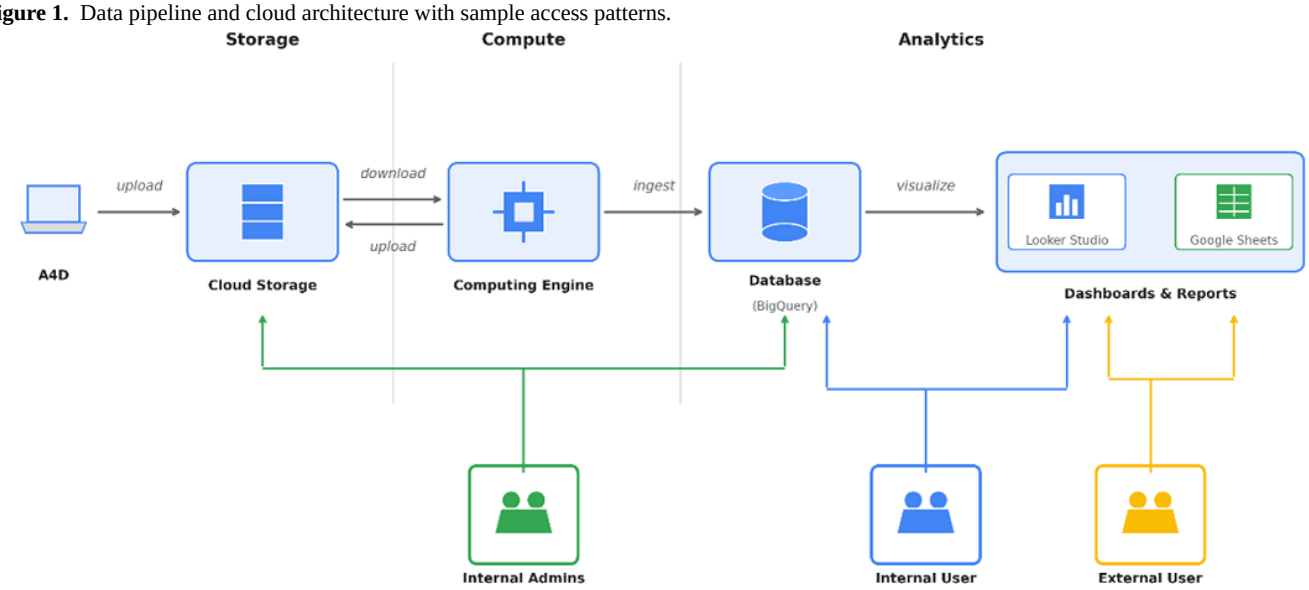
Report

As part of the preprocessing data pipeline, Action4Diabetes has been exchanging patient data with local program hospitals monthly via Microsoft Excel (Microsoft Corp) files [8]. The pipeline extracts the patient and medical product data in a largely standardized and unified manner, rendering it suitable for database storage and enabling consistent regional analysis; privacy compliance is embedded through anonymization protocols prior to upload, mitigating legal and ethical risks, and automated preprocessing and error reporting and tracking mechanisms largely replace manual data cleaning, improving accuracy and reducing delays. Data collected are subsequently uploaded to secure storage in the public cloud, where it is

XSL•FO
RenderX

processed and stored in a centralized database. Figure 1 illustrates the data pipeline and cloud architecture with sample access patterns.

Figure 1. Data pipeline and cloud architecture with sample access patterns.



We implemented the pipeline as an R package following a three-stage extract-clean-load architecture, with all source code version-controlled on GitHub. In the extraction stage, the pipeline reads monthly Excel worksheets from each hospital's tracker file, identifies data boundaries using marker rows and header patterns, and maps column names to a canonical schema through a configuration-driven synonym file. This step is necessary because column naming conventions vary considerably across hospitals and have changed over the 2017 - 2025 period; the synonym file currently maps over 60 patient variables and 14 product variables to their standardized equivalents.

The cleaning stage applies type conversions to every field and catches conversion errors through exception handling. Rather than silently dropping problematic records, the pipeline substitutes sentinel error values (eg, 999999 for numeric fields and 9999-12-31 for date fields) so that downstream analysts can trace exactly where data quality issues originate. Date parsing deserves particular mention because the trackers contain dates in many forms—Excel serial numbers, "15-Mar-2022," "Mar 2022," and occasionally dates embedded inside measurement strings—and the pipeline attempts each format in turn. For clinical measurements, the pipeline applies domain-specific rules: it strips inequality symbols from glycated hemoglobin values while retaining a threshold-exceeded flag, splits combined blood pressure strings into systolic and diastolic components, and converts height from centimeters to meters where the recorded value exceeds 50. Range validation then enforces clinically plausible bounds (eg, age 0 - 100 years, glycated hemoglobin 0% - 25%) and replaces outliers with error sentinels. Categorical fields such as insulin type and patient status pass through allowed-value checks defined in a YAML configuration file, and the pipeline logs every validation event as structured JSON with a standardized warning code, which makes it straightforward to audit data quality across sites and reporting periods. In the final stage, the pipeline assembles cleaned records into database-ready tables, selecting the relevant columns and ordering by patient and date for ingestion.

Anonymization takes place before any of the above processing begins. A separate Python tool reads the "Patient List" sheet from each tracker to build a lookup between patient names and unique program identifiers, which follow a structured format: a two-letter country code, a two-letter clinic code, and a three-digit sequential number. The tool then walks through every worksheet in the workbook and replaces each name with its corresponding identifier, saving the deidentified copy to a separate directory. Original files never leave secure local storage. During pipeline processing, the R code validates each identifier against the expected pattern and logs any mismatches for manual review. Once cleaning is complete, the pipeline serializes data to Apache Parquet and uploads it to Google Cloud Storage, from where it feeds into Google BigQuery—a managed columnar data warehouse that serves as the single point of access for analysis. Access control follows Google Cloud Identity and Access Management, with role-based permissions for internal administrators, standard users, and external collaborators (Figure 1); all storage buckets are private and restricted to authenticated project members.

The process of developing a data pipeline that ingests historical data and makes it accessible in a centralized database in a single place for further analysis took approximately 1 year to complete. The subsequent process stages entailed the identification of initial use cases that build upon the data, the implementation of dashboards and reports that facilitate data access, the automation of the data pipeline to ingest newly arriving data without manual intervention, and finally the establishment of a data governance strategy to oversee data content, structure, usage, and safety. Currently, the cloud database provides access to all historical data for the years 2017 - 2025 inclusive, encompasses 7 countries, 42 clinics, and approximately 1400 patients. Furthermore, the database contains information about the distribution and usage of approximately 200 medical products.

XSL•FO
RenderX

The impact of this work has been well documented. For example, as of August 16, 2021, 45 children, adolescents, and young adults with type 1 diabetes who were enrolled in Action4Diabetes's clinic support program constituted the first known cohort of Laotians to have survived a diagnosis of type 1 diabetes [5]. Given that as recently as 2016, no child was known to have lived with type 1 diabetes into adulthood in the Lao People's Democratic Republic, this is indeed a remarkable achievement for a lower-middle income country in Southeast Asia.

Despite its innovations, however, the Action4Diabetes pipeline faces several challenges that impact program efficiency and evaluation. Such challenges include fragmented data sources, creating inconsistencies and requiring extensive manual harmonization, and limited automation in data validation, meaning some errors still require manual correction. Further, the pipeline depends on Excel-based inputs, which restricts real-time integration and increases formatting inconsistency risk, while infrastructure gaps in partner hospitals—such as reliance on paper records—limit data completeness and accuracy, a problem also evident in high-income countries [9]. Finally, predictive analytics capabilities are basic, focusing mainly on supply forecasting rather than advanced patient risk stratification, and governance and interoperability challenges persist across multiple countries, with varying regulations and data-sharing agreements hindering streamlined data sharing and governance. These limitations highlight the need for further investment in automation, mobile data capture, and regional governance frameworks.

To enhance the effectiveness and sustainability of the Action4Diabetes data pipeline, several improvements are recommended. Transitioning from manual file uploads to secure application programming interface–based data exchange would reduce errors and improve timeliness. Expanding real-time dashboards for hospitals and health authorities could enable dynamic monitoring of patient trends and insulin stock levels. Integrating mobile data capture tools for frontline staff would minimize reliance on paper records and improve data completeness. Additionally, incorporating advanced predictive analytics such as machine learning models could support demand forecasting and early identification of high-risk patients [10]. Finally, formalizing regional governance frameworks and data-sharing agreements would ensure compliance, interoperability, and streamlined collaboration across multiple countries.

Limitations

The data pipeline and resulting dataset have several limitations that users of the data should bear in mind. Because the Excel tracker templates have evolved over the period 2017‑2025—for instance, individual insulin type checkboxes only appeared from 2024 onward—earlier records systematically lack variables that are present in more recent data. This creates a temporal bias: longitudinal analyses will inevitably draw on richer information for recent cohorts than for historical ones. Data completeness also varies across hospitals, largely as a function of local clinical capacity and staff familiarity with the tracker. In practice, this means that facilities with more established data entry routines contribute more complete records, and any analysis requiring complete cases will skew toward these better-resourced sites and the populations they serve.

Data fragmentation compounds the problem. Each hospital maintains its own tracker with locally adapted column names and entry conventions. The synonym-based harmonization catches most of these variations, but any field that does not match a known synonym is quietly excluded, so site-specific clinical observations can be lost. Manual data entry by clinicians—who understandably prioritize patient care over spreadsheet accuracy—introduces missing values, typographical errors, and inconsistent use of categorical codes (eg, free-text insulin regimen descriptions rather than the standardized categories the pipeline expects). The validation framework flags and logs these issues rather than silently discarding data, but the sentinel error values it substitutes must be handled carefully in any downstream analysis to avoid distorting summary statistics.

There is also a question of representativeness. The dataset captures only those patients enrolled in the Action4Diabetes program at participating hospitals; individuals with type 1 diabetes who are managed elsewhere or who have not been enrolled do not appear. In countries where program coverage is still limited, the data therefore cannot speak for the broader type 1 diabetes population. The monthly reporting cadence introduces a further constraint: clinical events that occur between reporting windows may be recorded with reduced temporal precision or missed altogether if not entered retrospectively. Finally, the anonymization approach—a deterministic name-to-identifier substitution—depends on the accuracy of each hospital's "Patient List" reference sheet. If a name is misspelled or absent from that list, the corresponding records will not be fully deidentified; the tool logs these cases for manual follow-up, but they represent a residual privacy risk that we actively monitor. Despite these constraints, the pipeline offers a pragmatic and reproducible way to harmonize messy real-world clinical data across a multicountry program, and its structured logging makes the boundaries of data quality transparent to anyone working with the resulting dataset.

Conclusion

The implementation of a centralized database solution for diabetes-related health care data is a challenging but also achievable objective when approached in a cost-effective manner, particularly when using services from modern cloud providers. With the current diabetes epidemic and projections for future decades illustrating a worsening state [11], the significance of data will likely continue to expand, enabling nongovernmental organizations to more effectively address the needs of communities they serve. The model used by the nongovernmental organization Action4Diabetes across Southeast Asia shows that much can be achieved and can perhaps be utilized elsewhere.

XSL•FO
RenderX

Acknowledgments

This work was presented as a poster at Advanced Technologies and Treatments for Diabetes (ATTD) Asia 2024, Singapore.

Funding

The authors declared no financial support was received for this work.

Data Availability

Not available due to ethical and consent considerations.

Authors' Contributions

SMN and MA conceived the manuscript. All authors cowrote and approved the manuscript.

Conflicts of Interest

SMN has received speaker fees from Insulet and Sanofi. The other authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

Multimedia Appendix 1
Action4Diabetes.
[DOCX File, 16 KB - diabetes_v11i1e89357_app1.docx]

References

1. Gregg EW, Patorno E, Karter AJ, et al. Use of real-world data in population science to improve the prevention and care of diabetes-related outcomes. Diabetes Care 2023 Jul 1;46(7):1316-1326. [doi: 10.2337/dc22-1438] [Medline: 37339346]
2. Nikita NA, Azim KS, Jafor AHM, Shayed AU, Hossain MA, Khan OU. Digital transformation in non-profit organizations: strategies, challenges, and successes. AIJMR 2024;2(5). [doi: 10.62127/aijmr.2024.v02i05.1097]
3. Toomey C, Gore J, Ooi F, et al. Action4Diabetes: a UK charity revolutionising type 1 diabetes healthcare across South-East Asia. Diabetes Care for Children and Young People 2021;1(3) [FREE Full text]
4. Ng SM. Action4Diabetes: a non‐profit organisation bridging the type 1 diabetes gap in Southeast Asia. Practical Diabetes 2023 Jul;40(4):6-10. [doi: 10.1002/pdi.2463]
5. Lek N, Manivong A, Rassavong K, Phommachack D, Toomey C, Ng SM. Type 1 diabetes in Laos, 2016-2021. Pediatr Diabetes 2022 Sep;23(6):620-626. [doi: 10.1111/pedi.13366] [Medline: 35598163]
6. Ng SM, Malene IV, Nguyen TK, et al. Internet analytics of an innovative digital educational resource of type 1 diabetes HelloType1 in local languages for people living with diabetes families and healthcare professionals in Southeast Asia. BMC Endocr Disord 2023 Nov 16;23(1):249. [doi: 10.1186/s12902-023-01501-4] [Medline: 37974071]
7. Wir bringen daten in die zivilgesellschaft!. CorrelAid ev. URL: https://correlaid.org [accessed 2025-09-10]
8. Ng SM, Malene IV, Nilar M, et al. Closing the Type 1 Diabetes gap in South-East Asia through government partnership working with non-government organisations. Diabetes Res Clin Pract 2022 May;187:109868. [doi: 10.1016/j.diabres.2022.109868] [Medline: 35395247]
9. Best J. NHS still reliant on paper patient notes and drug charts despite electronic upgrades, The BMJ finds. BMJ 2023 Sep 13;382:2050. [doi: 10.1136/bmj.p2050] [Medline: 37704231]
10. Lam JYJ, Barras M, Scott IA, Masood H, Abdel-Hafez A, Falconer N. Machine learning risk prediction models for medication harm in hospitalised adult patients. Ther Adv Drug Saf 2026;17:20420986251409325. [doi: 10.1177/20420986251409325] [Medline: 41567345]
11. IDF diabetes atlas 2025. : International Diabetes Federation URL: https://diabetesatlas.org/resources/idf-diabetes-atlas-2025 [accessed 2026-04-01]

Edited by S Li; submitted 11.Dec.2025; peer-reviewed by H Wu, J Ma; revised version received 08.Mar.2026; accepted 08.Mar.2026; published 17.Apr.2026.

Please cite as:
Ng SM, Aydinbas M, Martin T, Toh DJC, James S
Harnessing Digital Innovation for Diabetes Care: Insights From the Action4Diabetes–CorrelAid Data4Good Collaboration
JMIR Diabetes 2026;11:e89357
URL: https://diabetes.jmir.org/2026/1/e89357
doi:10.2196/89357

XSL•FO
RenderX

© Sze May Ng, Michael Aydinbas, Tyla Martin, Darren Jia Chen Toh, Steven James. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 17.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

T1D REACHOUT–A Mobile App to Deliver Peer-Led Mental Health Support to Adults Living With Type 1 Diabetes: Co-Design and Development Process

Sadie Lee[1], BA; Baray Sidhu[2], BSc; Parteek Johal[3], BSc; Ayman Azhar[4], BSc, BCS; Jonath Sujan[5], MEng; Matthias Görges[5,6*], PhD; Tricia S Tang[7*], PhD, RPsych

[1]Faculty of Arts, University of British Columbia, Vancouver, BC, Canada
[2]Faculty of Science, University of British Columbia, Vancouver, BC, Canada
[3]Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[4]Department of Computer Science, University of British Columbia, Vancouver, BC, Canada
[5]AI and Data Science Unit, Research Institute, BC Children's Hospital, Rm V3-324, 950 West 28th Avenue, Vancouver, BC, Canada
[6]Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada
[7]Division of Endocrinology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[*]these authors contributed equally

Corresponding Author:
Matthias Görges, PhD
AI and Data Science Unit, Research Institute, BC Children's Hospital, Rm V3-324, 950 West 28th Avenue, Vancouver, BC, Canada

Abstract

For adults living with type 1 diabetes (T1D), diabetes-specific mental health support is limited. Peer support and digital health platforms are promising strategies for delivering this support to this population, particularly those from geographically marginalized communities. Mobile apps, in particular, can enhance self-management and deliver support. This paper describes the iterative co-design and development process of a novel mobile app for use in a pilot trial, T1D REACHOUT (REACHOUT), that aims to reduce the diabetes distress, a core facet of diabetes-specific mental health, of adults with T1D. An initial think tank and 6 focus groups were conducted with adults with T1D to better understand their support needs and identify platform requirements. Following this, we partnered with adults living with T1D, the "end users," to iteratively co-design the REACHOUT app, enhancing usability and ensuring relevance. Adapting the open-source Rocket.Chat platform to user-defined requirements, we deployed the app in a single cohort pilot study. A network analysis of messages exchanged during the pilot study was performed to explore trends and patterns and demonstrate implementation feasibility. Pilot study outcomes informed further refinement before implementation in a randomized controlled trial. The implementation of the REACHOUT app features 6 key components identified in 6 initial focus groups: a 24/7 chatroom (a customized group messaging function with threads), topic-specific discussion boards, a peer supporter library, peer supporter profiles for a user-driven matching process, small group virtual sessions, and direct (one-to-one) messaging. Forty-six participants were encouraged to use any or all of the features as frequently as desired over a 6-month period during the pilot study. During this time, 179 private small groups were created, and 10,410 messages were sent, including 1389 chat room messages and 7116 direct messages; among these were 3446 messages exchanged between participants and their self-selected peer supporters. Key factors for successful implementation included (1) the co-design process involving comprehensive user engagement and (2) the opportunities realized through hybrid development. These findings offer generalizable lessons for mobile health research teams developing similar app-based interventions.

(JMIR Diabetes 2026;11:e71733) doi:10.2196/71733

KEYWORDS

peer support; type 1 diabetes; digital interventions; user-centered design; mobile phone; mobile health; mHealth; app development

Introduction

Mobile Apps for Type 1 Diabetes Management

In the era of digital health, mobile apps have increasingly been used to improve diabetes care. These platforms are vital for individuals living in geographically marginalized settings with limited access to conventional, in-person diabetes support [1]. Particularly, adults with type 1 diabetes (T1D) living in rural and remote communities in British Columbia have reported that peers with T1D share first-hand lived experience [1,2], from which they may benefit. As such, digitally delivered mental

XSL•FO
RenderX

health support may be a promising solution for underserved populations.

Several systematic reviews and meta-analyses involving children or adults with T1D have examined mobile apps [3-7] designed to improve self-management [8-10], monitor treatment through data visualization [11], and deliver psychosocial interventions [12-15]. Notably, research in this area has mainly focused on mobile app evaluation, while descriptions of app development lag behind.

Benefits of Involving End Users in App Development

Digital platforms that engage end users early in the development process are more likely to produce favorable results [16]. Specifically, co-design is a research methodology that, in principle, involves shared decision-making through the collaborative engagement of end users, researchers, and health care professionals to create and develop the technology to be used [17,18]. Thus, research exploring the mobile app preferences and perspectives of individuals with T1D has grown steadily [1,19,20]. Not surprisingly, 2 of the 4 best practices in app development involve "prioritizing user preferences" and "engaging stakeholders" [21].

While there is increased interest in collecting user feedback, how these data are integrated during the app development process has not been clearly described. Although several other T1D-specific apps provide detailed descriptions of their development, their goals are to improve glycated hemoglobin and blood glucose management by promoting self-care practices (carbohydrate counting, injection site rotation, insulin dosing, physical activity, etc) [22-25]. To our knowledge, no studies describe the comprehensive development process for a mobile app focused on reducing diabetes distress by offering peer-led mental health support to adults with T1D.

The T1D REACHOUT Intervention

T1D REACHOUT (REACHOUT) is a mobile app designed to reduce diabetes distress by offering peer-led mental health support to adults with T1D living in rural and remote areas of Interior British Columbia [2]. Based on self-determination theory, which highlights what motivates patients to make support-related decisions [26], REACHOUT is a 6-month intervention that offers a range of support delivery modalities to meet users' individual needs.

The use of the app is not intended to replace standard health care provision but rather to enable increased access to peer-led support and fill a gap in the health care system. While interdisciplinary diabetes care typically involves a team of endocrinologists, diabetes nurses, and dietitians, what is often overlooked and undervalued are the peers who understand first-hand the day-to-day experience of living with T1D.

Aims

This paper describes the iterative co-design and development process of the REACHOUT mobile app, including its use in a pilot trial and subsequent revisions made prior to its implementation in a randomized controlled trial (RCT). While empirical outcomes from the REACHOUT pilot intervention have been reported elsewhere [2], we focus specifically on the technical development and implementation of the app used in the intervention and aim to share generalizable lessons with other mobile health (mHealth) research teams.

Specifically, this paper reflects on the (1) collaboration process between researchers and end users (ie, adults with T1D) to identify information needs and support requirements that informed app functionality, (2) iterative design of the app with end users to enhance usability and meet key user needs, (3) deployment of the app in a pilot study to understand feasibility, and (4) incorporation of feedback from the pilot study to further refine the app before launching the RCT.

This paper is organized in the following sections: Ethical Considerations; Recruitment; Platform Design and Implementation, which includes all stages of the co-design process; Platform Features, which outlines the functionality of the designed app; Initial Platform Utilization Patterns, which includes a network analysis of app usage during the pilot study; Insights and Observations, which specifically addresses challenges to co-design in mHealth; and Discussion, which includes an overview of our findings, limitations, and future work.

Ethical Considerations

The University of British Columbia Behavioural Research Ethics Board approved this project (H20-00276, PI Tang, date of first approval 2020-03-31), with additional harmonized review by the Interior Health Research Ethics Board (same approval number). Participants were given study information, allowed to ask questions, and provided electronic consent. Focus group participants received a US $50 gift card upon completion. Pilot study participants received a US $25 and a US $50 Amazon gift card at baseline and at 6 months, respectively. Focus group discussions throughout the co-design process were recorded, transcribed verbatim, and then de-identified. Study data were stored securely with access limited to study team members.

Recruitment

For our preliminary think tank session, we recruited participants (n=17) from a list of adults with T1D who expressed interest in T1D-specific mental health support at a Diabetes Canada-sponsored T1D patient advocacy event.

Focus group participants (n=38) were recruited using several modalities: (1) inviting participants from previous T1D diabetes studies, (2) obtaining referrals from health care providers, (3) sending emails to members of Breakthrough T1D living in the Interior British Columbia regions, (4) posting messages on T1D-specific social media sites, and (5) word of mouth [1]. Eligibility criteria included being diagnosed with T1D, ≥19 years, living in the Interior Health Authority region, and able to speak English [1].

Pilot study participants (n=46) were also identified and recruited via (1) diabetes education centers in the Interior Health Authority region, (2) T1D advocacy organizations, (3) T1D-specific Facebook support groups and social media outlets, (4) referrals from health care providers, and (5) word of mouth

XSL•FO
RenderX

[2]. Eligibility criteria were similar to those of the focus group study, except that participants also needed to have access to the internet and/or a smartphone [2].

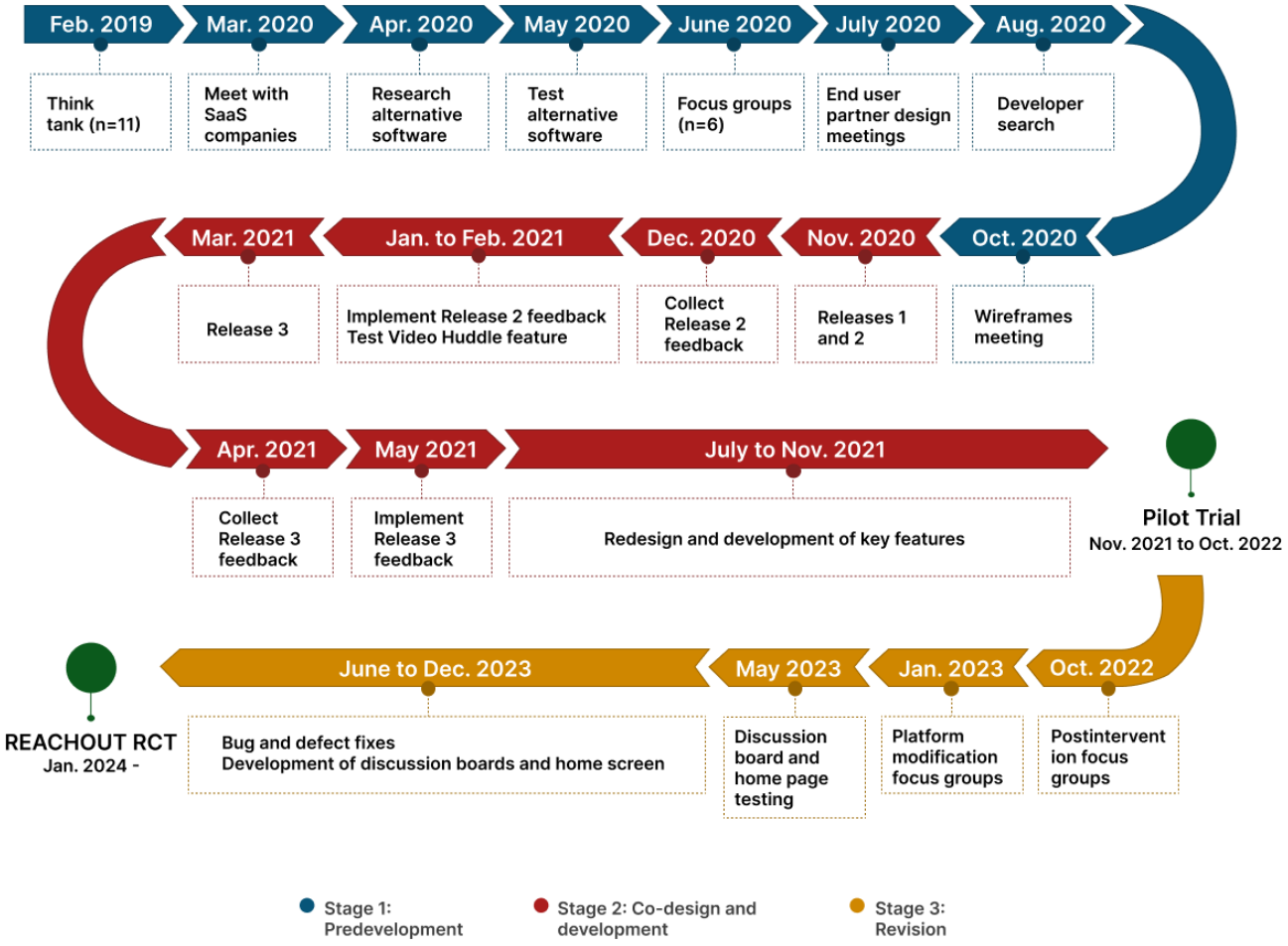
End user partners (n=11) were adults living with T1D who contributed to the iterative development process and were recruited from participants in the previous focus groups.

Platform Design and Implementation

Overview

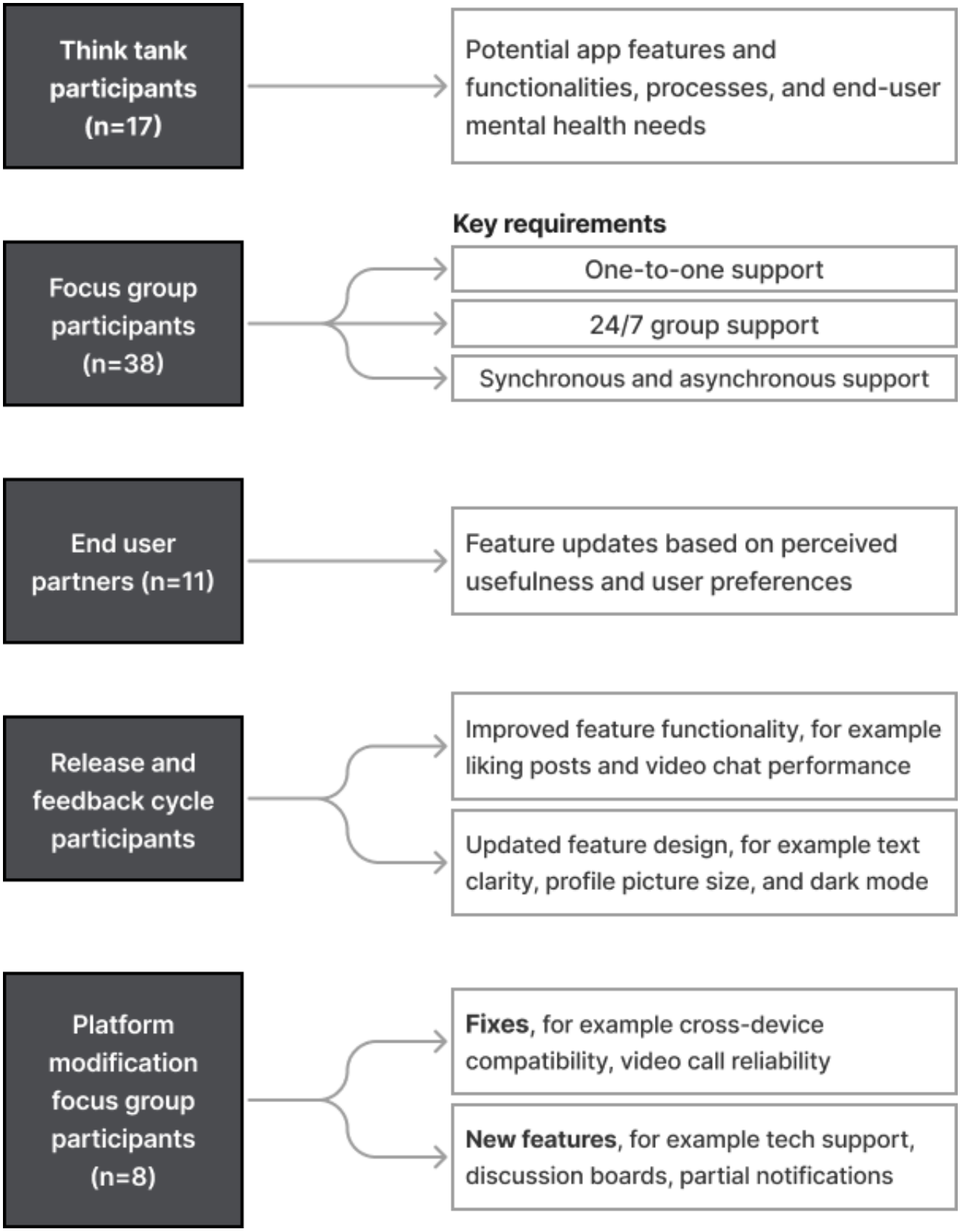
Our development process occurred in 3 stages: predevelopment, co-design and development, and revision (Figure 1). As people with T1D are the end users of the REACHOUT app, their involvement was imperative; hence, their feedback and guidance were sought throughout the app's development process to ensure it met end user needs. Through focus groups, informal think tanks, and feedback sessions, the co-design approach defined requirements and critical features for the peer support–based mobile app (Figure 2).

Figure 1. Timeline of the REACHOUT mobile app development process. Stage 1: predevelopment is indicated by blue arrows; stage 2: co-design and development are indicated by red arrows; and stage 3: deployment and revision are indicated by yellow arrows. Study milestones, including the pilot trial and the randomized controlled trial (RCT), are indicated by green circles. SaaS: software as a service.



XSL•FO
RenderX

Figure 2. Co-design and development process, with each of the user groups engaged and the actions performed as a result of the activity.



Stage 1: Predevelopment

Think Tank

A preliminary think tank of adults with T1D was used as a scoping activity to elicit the support needs of this group and explore the possibility of a digital platform as the modality for emotional support. The session lasted approximately 60 minutes and was audio-recorded, with only an audio transcript available. Insights from this discussion informed the subsequent pursuit of a mobile app–based intervention and the continued use of a co-design approach.

XSL•FO
RenderX

Participants identified potential support delivery features that would be beneficial, including individual peer-to-peer support, group-based messaging, professional moderators, and discussion boards. Additionally, participants discussed considerations when choosing peer supporters, such as using the same technologies (eg, the same continuous glucose monitor).

Platform and Software Selection

Given the anticipated benefits of a flexible setup and the ability to customize the app, digital health software companies were first reviewed. These included Weltel, Latero Labs, Freshworks, Tactica Interactive, and Tech Samurais (currently Four Nine Digital). However, due to financial constraints prohibiting the development of a de novo app, we decided to develop the app by adapting and expanding existing communication–based software. Alternative solutions were researched to identify which would best fit the identified requirements while making only limited changes to the app.

Slack, Mattermost, and Rocket.Chat were identified as potential platforms (Table 1). Slack was disqualified due to a lack of options for modification given its closed-source software, leaving Mattermost and Rocket.Chat as the final candidates for consideration. Rocket.Chat was selected for its customization, comprehensive documentation, and polished user interface and user experience design. Subsequently, Rocket.Chat was tested with end user partners (n=11) to evaluate user experience before being modified to create the REACHOUT app.

Table . Alternative software options and criteria researched[a].

Property	Slack [27]	Rocket.Chat [28]	Mattermost [29]
Customizability	Slightly customizable	Very customizable	Very customizable
Pricing (per user)	CAD $207/year; US $ ~145/year	Free if self-hosted or CAD $55/year; US $ ~38/year	Free if self-hosted or CAD $141/year; US $ ~99/year
Hosting costs	Included	CAD $2045/year; US $ ~1430/year	≤CAD $2100/year; US $ ~1469/year
Setup and maintenance	Easy and nontechnical	Complex	Complex
Hosting location	United States	Canada	Canada
Source	Closed-source	Open-source	Open-source

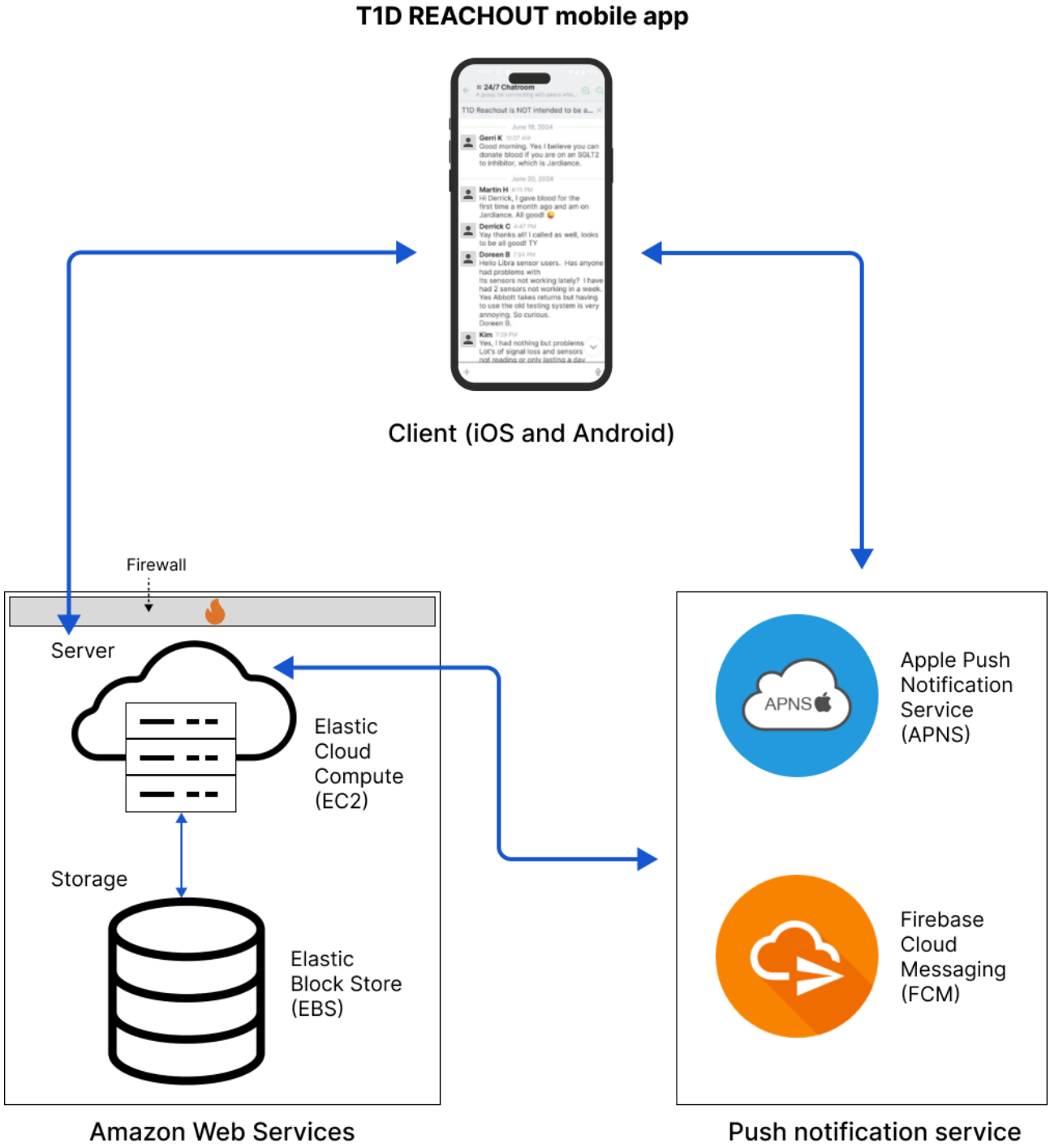
[a]Pricing in Canadian dollars reflects amounts at the time of inquiry. Conversion to US dollars was done on January 20, 2025.

Technology Stack and Platform Architecture

REACHOUT is built on the Rocket.Chat cross-platform messaging platform, powered by React Native. When new updates are made to the app, we build iOS and Android app files from the latest code changes using Xcode and Android Studio. The client-facing component includes an app for Android and iOS devices that interacts with our Amazon Web Services (AWS) backend via a firewall. An Amazon Elastic Cloud Compute instance provides compute capacity with Amazon Elastic Block Storage for long-term data storage. Initially hosted on DigitalOcean's cloud, it was later migrated to a Canadian AWS instance at the recommendation of initial contractors and supported by more robust documentation from Rocket.Chat for AWS deployment. Push notifications use Apple Push Notification Service for iOS devices and Firebase Cloud Messaging for Android devices (Figure 3).

XSL•FO
RenderX

Figure 3. Overview of the current app architecture with interactions between client components, server and storage components, and third-party notification services. T1D: type 1 diabetes.



Focus Groups

Six focus groups with 38 participants (5 - 7 participants per focus group) were held to identify users' current and past T1D-related mental health needs, explore desired features for the app, elicit peer supporter communication mode preferences, and understand patterns of social media use for diabetes-specific support [1]. Detailed results have been reported previously [1].

Through these focus groups, participants established the following key requirements:

1. One-to-one support from a peer supporter (an individual with similar firsthand experience) could help users navigate the "ups and downs" of living with T1D.
2. 24/7 group support allows participants to receive support at any time of day, including evenings and weekends. It is difficult to predict when users need immediate emotional validation, normalization, advice, and acceptance. For example, participants cited hypoglycemia, which can occur in the middle of the night, as a situation in which immediate emotional support can provide reassurance and comfort.
3. Synchronous and asynchronous support options are available for meeting virtually with others or asking or responding to questions in participants' own time.

XSL•FO
RenderX

Focus group participants also provided feedback on specific app features that would be beneficial, further informing development [1]. These recommendations included (1) which data to incorporate in peer supporter profiles, such as age of diagnosis, hobbies, stage of life, and T1D monitoring and treatment technologies used (eg, brand of continuous glucose monitor) so that participants can determine a suitable match; (2) having trained health professionals on the platform to prevent the exchange of medical information; (3) including resources and links to keep up with best practices; and (4) hosting topic-specific group discussions on subjects specific to living with T1D.

From focus group discussions, we identified 6 key features based on recurring needs and priorities articulated by participants and prioritized in development based on feasibility [1]: peer supporter profiles, a peer supporter library, one-to-one messaging, group messaging, topic-specific discussion boards, and virtual huddles.

Stage 2: Co-Design and Development

Release and Feedback Cycle

An external developer was contracted to translate the user-identified key features into an initial prototype. Three iterations of the prototype were developed and released to subsequent groups of test users. Over the course of each iteration, users were asked to provide feedback to the research team via Zoom meetings regarding feature design preferences, functionality, and pain points (eg, a lack of intuitiveness in performing an action).

The first iteration (released to 23 users: 17 iOS users and 5 Android users) included peer supporter profiles, one-to-one messaging, and group messaging, which were more feasible to implement given Rocket.Chat's existing functionality. Feedback was received and implemented in the following iteration. The second iteration (released to 43 users: 31 iOS users and 12 Android users) included the existing features and added video huddles. The third iteration retained the features we had already updated (peer supporter profiles, one-to-one messaging, and group messaging), updated the video huddles feature, and added a peer supporter library.

There was approximately 25% overlap in users across iterations, as users participated in consecutive app iterations. We sought to recruit 50 participants and recruited 46. Given that this was a feasibility study, no power calculation was required. To determine a reasonable sample size, we conducted a literature review and identified other pilot studies similar in topic. According to a systematic review of mHealth apps for chronic illness and/or social support, 12 of the 13 studies reported samples of fewer than 50 participants, with an age range of 12 to 49 years [30]. Considering these studies, we elected to recruit a sample size of 50 participants.

End User Partners

A small group of end user partners (n=11) was engaged to support iterative collaboration throughout development. End user partners contributed to wireframe development, design recommendations, and postimplementation feedback, such as ranking the importance of content on peer supporter profiles and rating the usefulness of proposed features. Feedback was elicited through 3 group meetings, a one-on-one phone call, and an email discussion. This enabled end users to influence design decisions and support iterative development grounded in lived experience.

The mean age of end user partners was 39 (SD 11) years; 55% (n=6) identified as female. The small number of end user partners was not intended to be representative of the T1D population as a whole but rather to support meaningful co-design through in-depth engagement.

Stage 3: Deployment and Revision

Deployment

Following these 3 development iterations, the app was tested in a pilot study published elsewhere [2] and involved 46 users receiving support and 36 peer supporters delivering support. The mean age of users receiving support was 44 (SD 15) years, with the majority (n=29, 76%) identifying as women. The mean age of peer supporters was 41 (SD 16) years, with 37 (73%) identifying as women [12]. Of the 46 participants onboarded during app deployment, 8 withdrew for personal reasons. Of the 37 peer supporters onboarded, 2 withdrew, also for personal reasons.

We determined which participants were iOS or Android users and collected their respective App Store and Google Play Store email addresses. For iOS users, we used TestFlight to distribute the app builds. Users were provided access to a test version of the app. When new versions of the app were released, they were either automatically updated on the users' mobile phones or could be downloaded manually via the TestFlight app. For Android users, we initially distributed an Android Package Kit file that users downloaded, but later pivoted to releases on the Google Play Store for ease of use. Users were sent a link to download the app, and any updates were automatically downloaded to their mobile phones. Only invited users could download the app for either Android or iOS.

Platform Modification Focus Groups

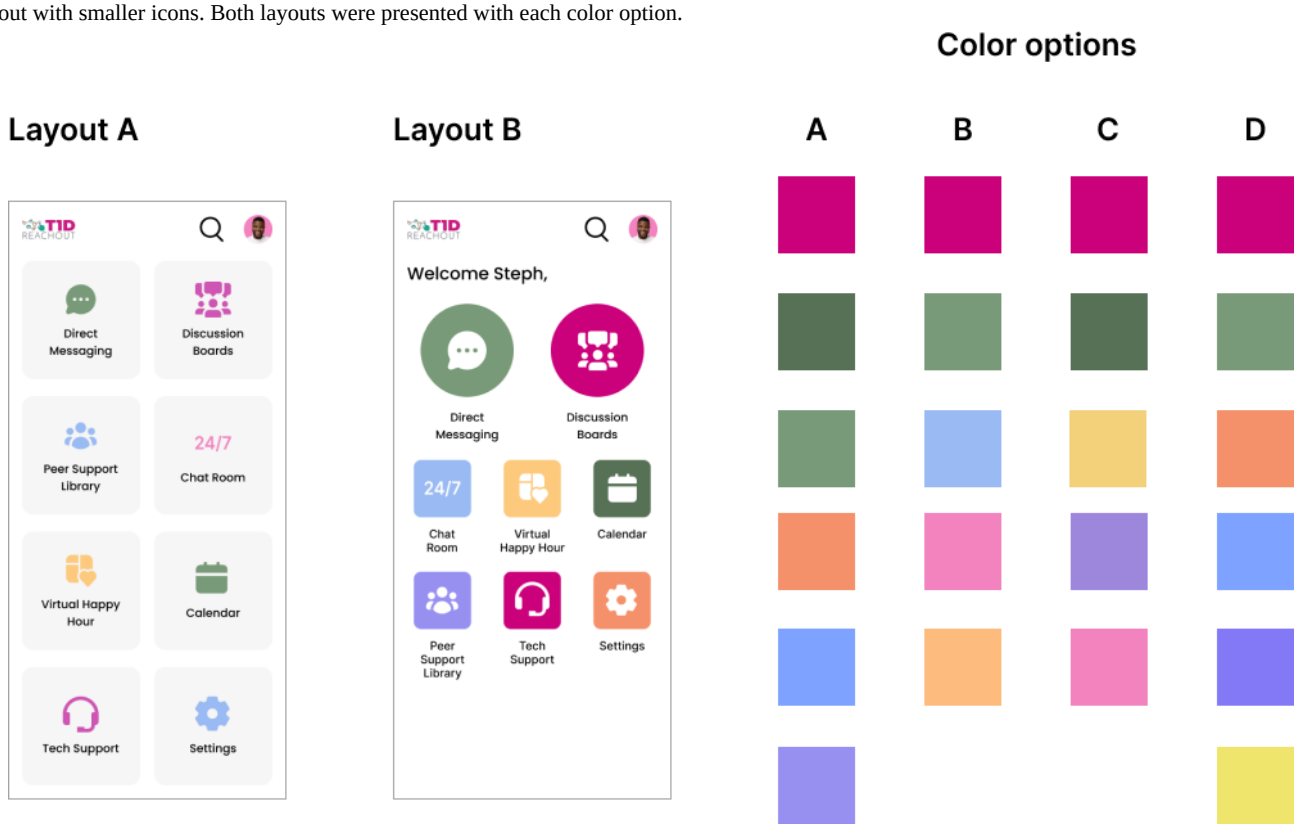
A further set of postpilot focus groups was held to receive feedback on user priorities for the app and potential modifications. Features that needed to be fixed and features to add based on user feedback were presented [31], and users (n=8) completed a prioritization exercise, classifying features into must have, should have, could have, and won't have [32].

User Testing of Discussion Board and Home Page

Given discussion boards as a high priority, user testing was conducted to determine user flows that support both anonymous and identifiable posts on a discussion board and to allow users to bookmark (save) posts for later reference. Feedback was also received on the design of the discussion board menu and the home page (Figure 4).

XSL•FO
RenderX

Figure 4. Home page user testing with 2 layout options and 4 color options. Layout A uses a card layout with larger icons, while layout B uses a grid layout with smaller icons. Both layouts were presented with each color option.



Further Postdeployment Development and Implementation

The new features and modifications identified through the postpilot focus groups and user testing were translated into wireframes by a contracted designer and then implemented by a contracted developer for the RCT of REACHOUT. Our platform team was required to fix bugs and defects internally and implement user-requested features that the developer had not delivered. These internal changes resulted in a 4-month delay in the launch of the REACHOUT app for the RCT.

Platform Features

Overview

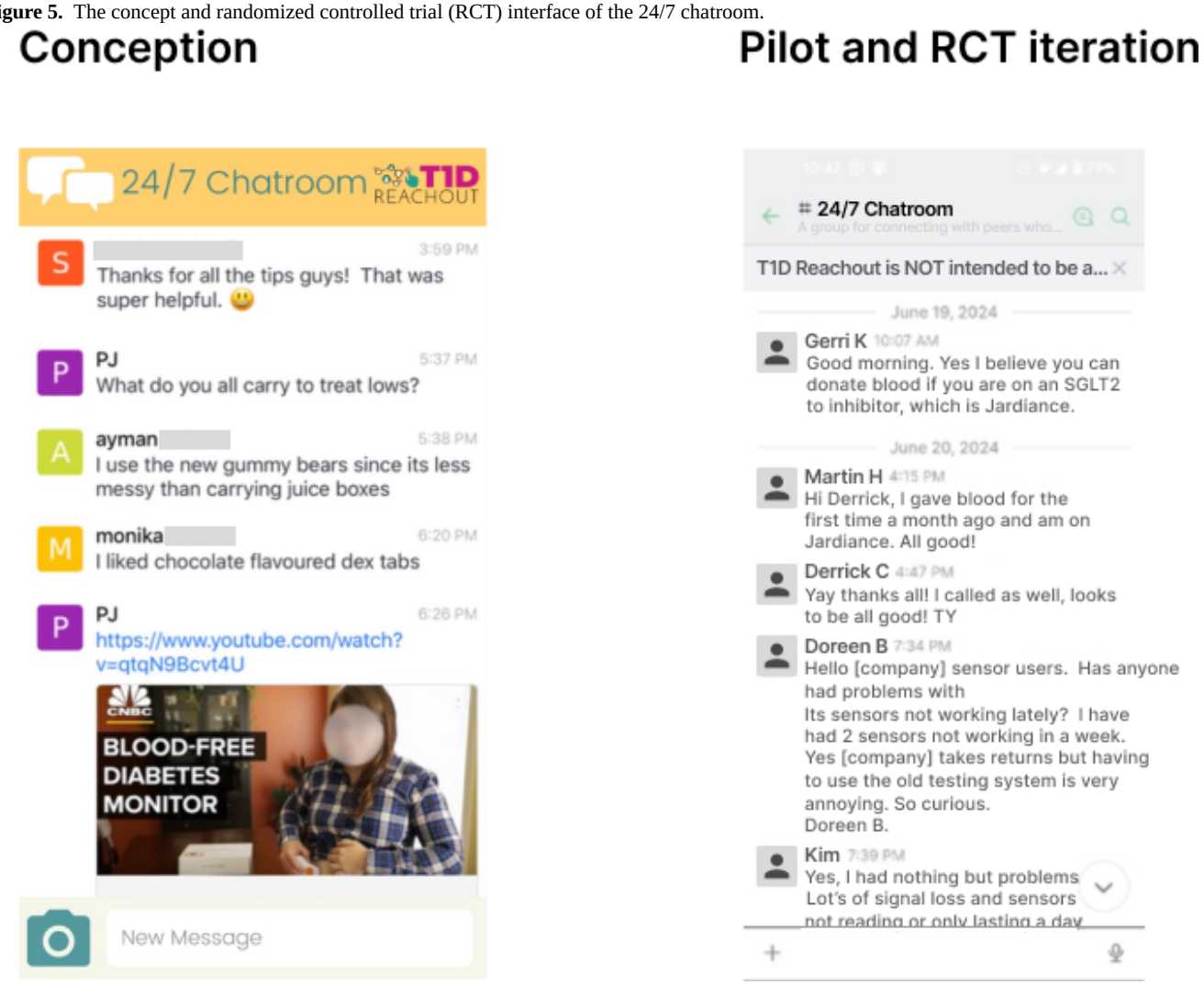
The iteration of the REACHOUT app that was released for the RCT contains the following 6 key features identified in the initial stage 1 focus groups: a 24/7 community chat room (group text messaging), topic-specific discussion boards, a peer supporter library, peer supporter profiles, small group virtual sessions (ie, virtual happy hours), and direct (one-to-one) messaging. Each of these features is described below, along with an accompanying figure showing the progression of development based on user feedback, from earliest design (conception) to the final iteration used for the RCT.

24/7 Community Chat Room

The 24/7 chat room is a feature built upon Rocket.Chat's existing microservices architecture (Figure 5). Users can access a general-purpose discussion forum and send text, images, and voice messages. A banner reminds users of community rules, such as not providing peers with medical advice. In fact, group exchanges are monitored by research assistants and health care professionals (registered psychologist, dietitian, diabetes nurse, and pharmacist).

XSL•FO
RenderX

Figure 5. The concept and randomized controlled trial (RCT) interface of the 24/7 chatroom.

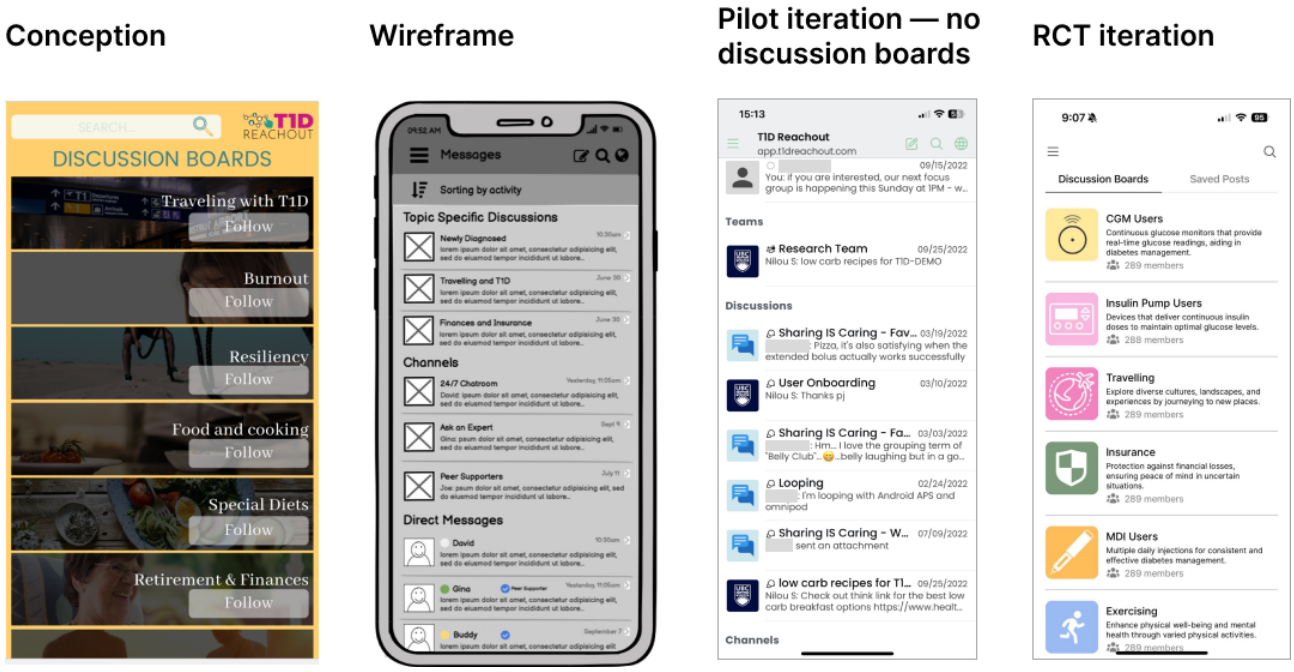


Topic-Specific Discussion Boards

According to pilot study feedback, exchanges on the 24/7 chat room were perceived as overwhelming to some users, and message threads could easily be lost [31]. This formed a new requirement: creating topic-specific discussion boards for focused messaging channels limited to a specific T1D topic. These discussion boards were developed as separate public messaging channels using Rocket.Chat's channel creation feature. The app was modified to create a designated menu page for topic-specific discussion boards. User-identified topics on current discussion boards include exercise, multiple daily injections, insulin pumps, insurance coverage, continuous glucose monitors, and traveling (Figure 6). The research team can add additional discussion boards, and frequently initiated topics will be retained for future platform instances.

XSL•FO
RenderX

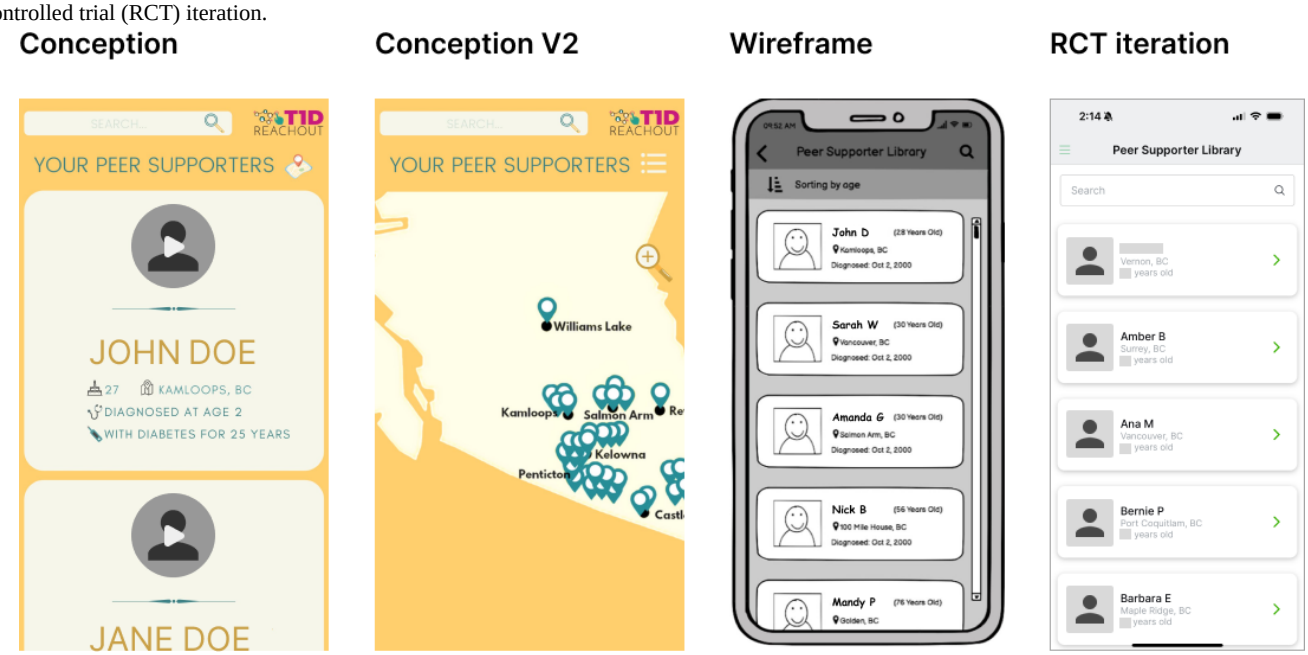
Figure 6. Discussion board iterations, including conception, wireframe, pilot iteration, and the randomized controlled trial (RCT) iteration, based on user feedback from the co-design and development process.



Peer Supporter Library

The peer supporter library invites participants to self-select a peer supporter [2]. The implementation of this feature was initially explored in two ways: (1) a participant (one who receives support) browses profiles from the library and then chooses their own peer supporter (active matching) or (2) a participant conveys their preferences to the research team, which assists with selection (passive matching). The active approach was adopted so that users could identify a peer supporter who best fits their unique support needs. For example, participants considering having children might prefer a peer supporter who is already a parent. Rocket.Chat's mobile app was customized to add this peer supporter library view (Figure 7). The library previews the first name and last initial, location, and age displayed as a card, with access to review the complete peer supporter profile enabled by tapping each profile card.

Figure 7. Peer supporter library iterations, including conception, a second conception with a location-based library, a wireframe, and the randomized controlled trial (RCT) iteration.



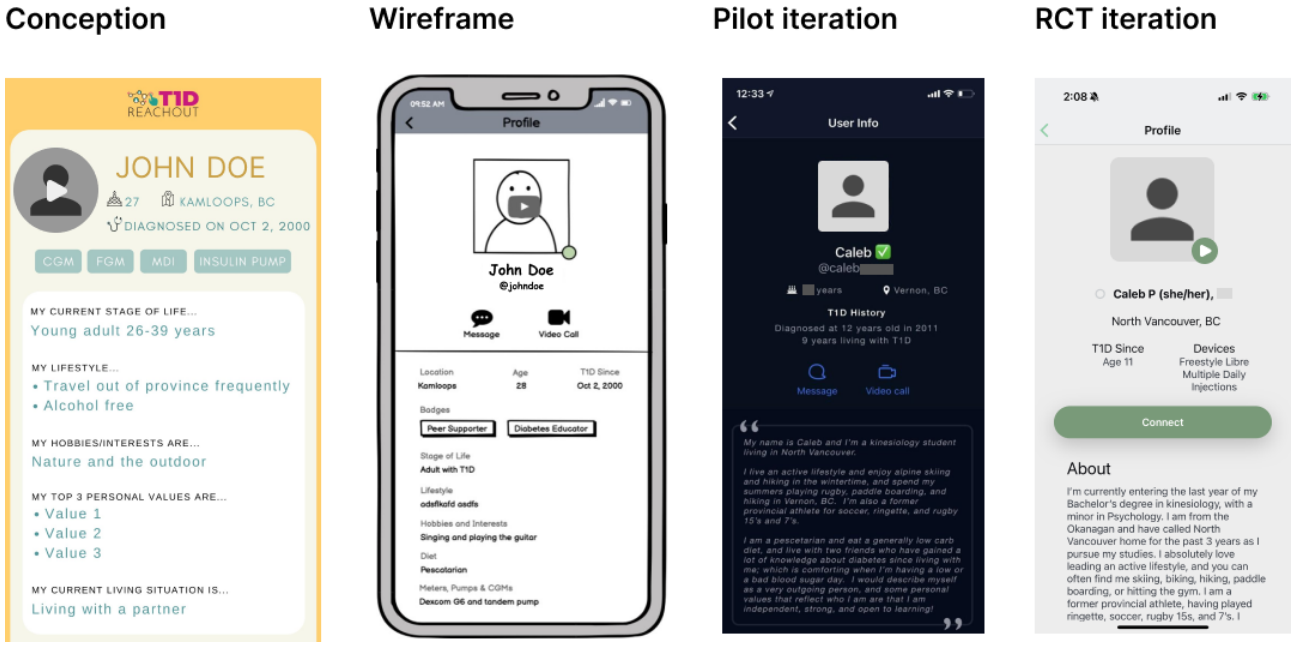
Peer Supporter Profile

To initiate an active matching approach, participants must determine which peer supporter best fits their support needs. To aid this, each peer supporter completes a profile that can be accessed through the peer supporter library [2]. The characteristics included in each profile (Figure 8) are first name and initial of last name, age, age at the time of diagnosis, and devices used (eg, type of insulin pump and continuous glucose monitor). Each profile also contains a written biography and a

XSL•FO
RenderX

1-minute video biography accessible via a private YouTube link [2].

Figure 8. Peer supporter profile iterations, including conception, wireframe, pilot iteration in dark mode, and the randomized controlled trial (RCT) iteration.



Small Group Virtual "Happy Hour" Sessions

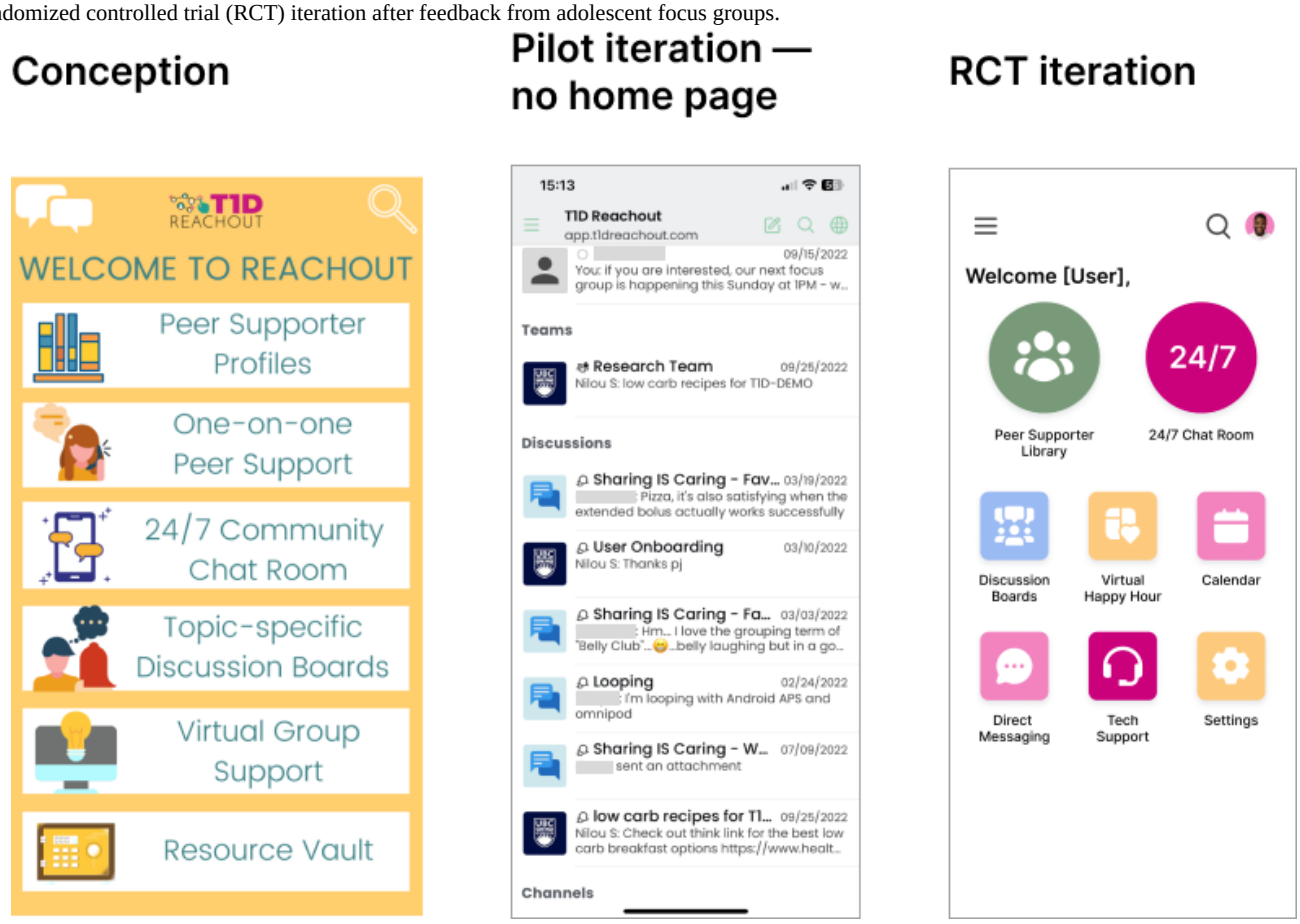
"Happy hours" are virtual synchronous small group meetings led by peer supporters. They highlight distress-related themes, such as workplace bias and discrimination. Currently, information on virtual happy hours is shared via announcements posted to a discussion board, and sessions are held on Zoom (Zoom Video Communications Inc.). For the RCT, virtual happy hours replaced virtual huddles, featured in the previous pilot study [2].

Home Page

Adolescent focus groups identified a new requirement [33]: an enhanced user interface and user experience, achieved by making the home page less intimidating and more visually appealing. One example was a landing page shown when the app opened (Figure 9). The home page also functions as a directory where users can navigate to other app areas.

XSL•FO
RenderX

Figure 9. Home page iterations, including conception, pilot iteration without a home page (which directly jumped into the 24/7 chat room), and the randomized controlled trial (RCT) iteration after feedback from adolescent focus groups.



Initial Platform Utilization Patterns

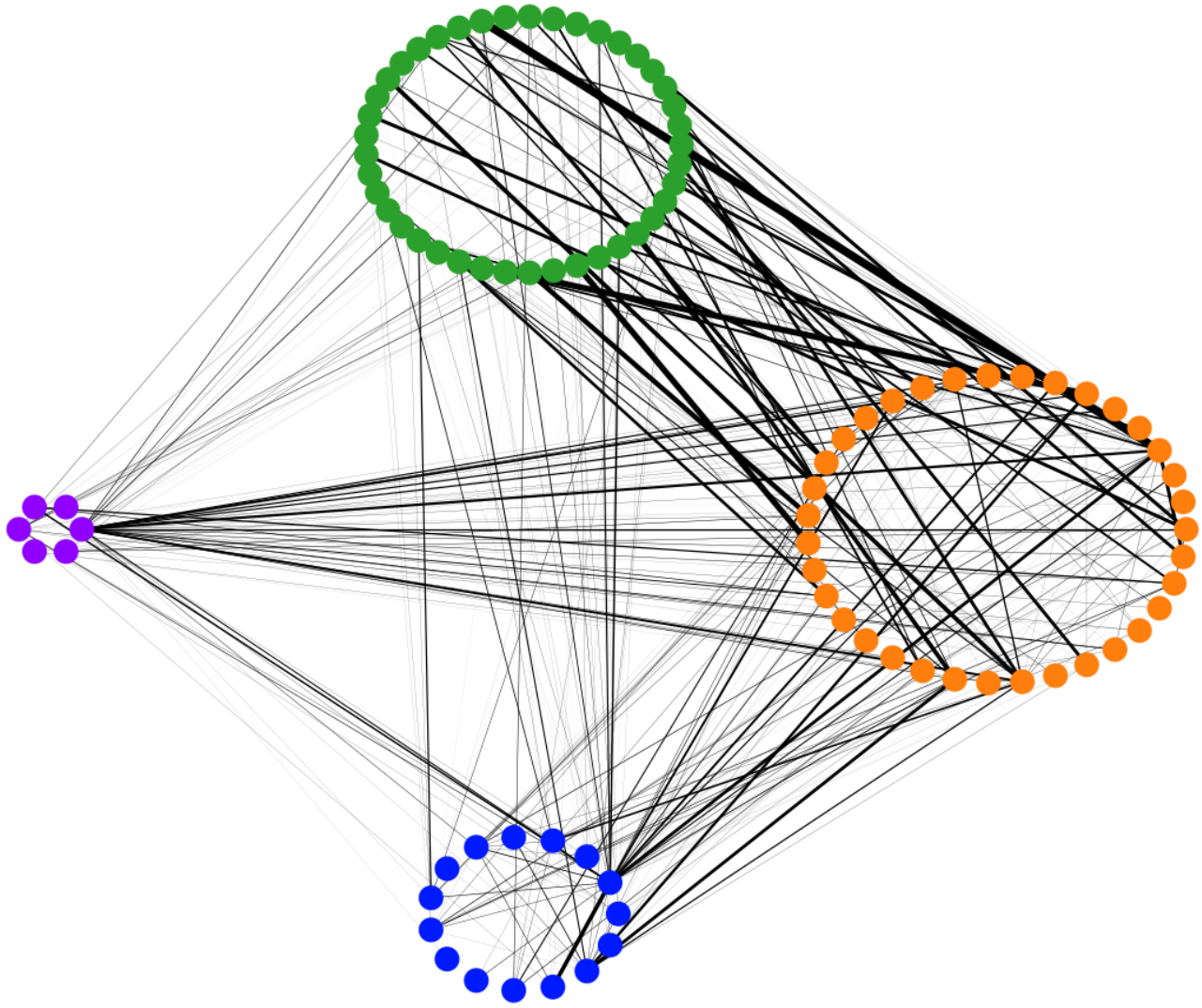
To analyze trends in this peer network intervention, timestamps of messages, along with their senders and recipients, were logged between October 12, 2021, and March 15, 2022. A network analysis was performed using NetworkX [34], and usage patterns were interpreted qualitatively by identifying patterns (numbers and widths) in edges, direction of communication, and temporal characteristics as the study progressed.

During a 6-month period when the app was deployed for the REACHOUT pilot study, 10,410 messages were sent, including 7116 one-to-one direct messages and 1389 chatroom messages; 179 private messaging groups were created. One-to-one messaging was the app's most common mode of written communication and was a key requirement. The most one-to-one messages were between peer supporter-participant pairs (3446), followed by 802 between peer supporters and health care team members. Users explained that group messaging, such as the 24/7 chatroom, was more commonly browsed than contributed to.

Substantial communication occurred between participants and their self-selected peer supporters (Figure 10), likely driven by an intervention requirement to engage with their assigned participants at least weekly (later changed to every 2 weeks based on feedback). This illustrates the app being used as intended, that is, as a mechanism to deliver peer-led mental health support primarily between a participant and their self-selected peer supporter. A total of 3446 messages were exchanged between participants and their self-selected peer supporters. However, communication methods varied among peer supporter-participant pairs, with some preferring Zoom or phone calls, which were not captured in the network graphs. Frequent communication also appeared to occur between peer supporters themselves and between peer supporters and the health care team.

XSL•FO
RenderX

Figure 10. Network analysis showing communication patterns among app users by user type, specifically the defined roles assigned to users within the app (blue: research team, purple: health care team, orange: peer supporters, and green: participants). The number of messages sent is represented by the line width (edge weight), while the directionality of messages is not considered in this graph.



Insights and Observations: Challenges to Co-Design in mHealth

Through an iterative co-design process, we learned the importance of end-to-end user engagement to improve the user experience and ensure users' needs are met. This collaboration refined the features developed from early iterations (eg, user-identified requirements for one-to-one support via direct messaging and group support in the 24/7 chatroom) into later iterations (eg, developing topic-specific discussion boards based on user feedback). Yet, the process was also challenging, particularly in establishing consensus on requirements across several focus groups, and when users' stated wants did not align with what they needed [35].

The focus groups used in the app design process varied in size—this meant that the needs of 1 group were not necessarily representative of all users' needs. We would seek input from different groups and, understandably, receive different opinions from each. It may have been beneficial to have a process to make the final call in these scenarios. We attempted strategies, such as voting or "buying" features, but ultimately, not everyone would be satisfied, as users will necessarily have unique needs.

While we had a general understanding of how features should function, the design and functional requirements were often not communicated clearly between the designer and the developer. Consequently, the delivered features sometimes differed from the research team's expectations. For example, the home page design was not implemented as displayed in the mock-up. We sought to resolve this by adding wireframes and gathering feedback from the team or participants earlier. However, frequently incorporating participant feedback to enhance the user experience sometimes led to revising features already developed or in progress, which hindered our progress.

Given that Rocket.Chat is a large open-source codebase with the React Native repository currently having 131 contributors, it took new developers significant time to onboard and become familiar with the codebase, which also led to delays. Furthermore, because our resources only allowed for a single contracted developer working part-time (20 hours per week), always remotely, and sometimes in different time zones, it took substantial effort to receive deliverables on time.

XSL•FO
RenderX

To mitigate these challenges, we took a hybrid approach following the pilot study for subsequent development and maintenance of the REACHOUT app by hiring a vendor to handle the development of complicated and time-consuming tasks, which were of lower priority, and an in-house product manager with technical expertise to manage the execution of the product development and budgetary management, proper communication of requirements to the vendor, biweekly release of the app for internal testing, and continuous assessment of the work performed by the vendor to resolve issues promptly. We also had internal student developers with a background in computer science on the research team that worked on urgent features, minor improvements, bug fixes, and internal app testing. This approach ensured improved coordination, task prioritization, project management, efficient development and quality control, and better documentation, helping eliminate skill gaps. Though we acknowledge that there are vendors who offer quality service at a reasonable cost and other vendors who are expensive, but take care with proper coordination, communication, management, design, development, and testing to ensure a quality product is delivered to their customers, we found our hybrid approach to be sustainable and budget-friendly compared to fully internal or fully external development, allowing us to release a stable and reliable version of the app to our users in a reasonable timeline.

Discussion

Overview of Findings

In collaboration with adults from the T1D community in British Columbia, we designed and implemented REACHOUT, a mobile app for adults with T1D in British Columbia, which seeks to reduce diabetes distress (a core facet of diabetes-specific mental health). This project builds upon previous work establishing virtual peer support as a viable strategy for reducing diabetes distress [1,13]. Six key user-defined features for the app were identified through focus groups and refined through co-design and a pilot deployment: peer supporter profiles, a peer supporter library, one-to-one messaging, group messaging, topic-specific discussion boards, and small group virtual sessions.

The REACHOUT app uniquely features a user-driven “match-making” process to access one-on-one peer support [2]. Specifically, each user chooses their peer supporter based on the factors that matter most to them (eg, shared hobbies, lifestyle, insulin pump and/or continuous glucose monitor brand, location of residence) rather than the selection determined by the research team or those administering the app. An iterative co-design approach enabled early and comprehensive user engagement, identified app requirements, enabled prompt identification of pain points (eg, usability issues and bugs), and provided us with ongoing formative feedback from end users.

Limitations

Our app development process had its limitations. First, no formal usability testing was undertaken because (1) the importance of such testing was not initially identified and (2) the tight timelines and constrained resources of our behavioral research study were challenging. Formal usability testing using standardized questionnaires, such as the System Usability Scale [36], Questionnaire for User Interaction Satisfaction [37], or Post-Study System Usability Questionnaire [38], could have ensured that users were sufficiently able to use the app and contribute to our overall understanding of feasibility. While the lack of formal usability testing may have affected the app’s quality, user feedback from focus groups was incorporated before the app was used in the RCT. Additionally, focus group participants were predominantly female and self-selected, which may not reflect the larger T1D community in British Columbia. Future work should ensure diverse voices are represented when establishing app requirements, for example, by increasing the number of male participants.

Future Work

The app has been deployed and is being evaluated in the REACHOUT RCT (NCT05668507). A formal usability evaluation should be performed to examine the app’s functionality and identify opportunities for additional features or device optimization. Features currently in development, based on further user feedback, include allowing peer supporters to limit the number of participants they connect with. When evaluating usability, participants’ usage practices, motivations, and intents should also be examined to understand ideal usage patterns. We are also codeveloping an adolescent-focused version of the app [33], which considers adolescents’ unique support needs and will be evaluated in a future trial.

Conclusions

The REACHOUT app was developed to reduce diabetes distress by providing peer-led mental health support to adults living with T1D in British Columbia. The app development process was described to guide others on similar projects. An iterative co-design process with end-to-end user engagement and continual user feedback was beneficial, enabling a thorough understanding of end user needs, preferences, and context, and ensuring the final app was impactful and relevant to end users. The REACHOUT app offers choice-based, just-in-time, and customizable support, making it the optimal peer support model for individuals of all socioeconomic backgrounds, geographic residences, and life circumstances [2]. Its potential to reduce diabetes distress and improve diabetes-specific mental health outcomes is being evaluated in a large-scale province-wide RCT.

Acknowledgments

The authors thank all participants for sharing their ideas and insights and providing constructive feedback to shape the T1D REACHOUT platform and Nicholas West for editorial assistance. This work was previously presented, in part, at the American

XSL•FO
RenderX

Diabetes Association's Annual Scientific Sessions in June 2022 in New Orleans, Louisiana, and June 2023 in San Diego, CA. Generative artificial intelligence tools were not used in this manuscript's writing, editing, or development.

Funding

This study was supported, in part, by research grants from the Breakthrough T1D (2-SRA-2020-986-S-B), Breakthrough T1D Canada and Brain Canada (2- SRA-2022-1183-S-N). TST is supported by a Michael Smith Health Research BC/Breakthrough T1D Canada Health Professional-Investigator Award (HPI2021-2359). MG was supported by a Michael Smith Health Research BC scholar award (SCH-2020-0494) for the duration of the study.

Data Availability

The authors cannot share data due to privacy considerations agreed to by the research ethics board.

Conflicts of Interest

None declared.

References

1. Ashrafi S, Taylor D, Tang TS. Moving beyond "don't ask, don't tell": mental health needs of adults with type 1 diabetes in rural and remote regions of British Columbia. Diabet Med 2021 May;38(5):e14534. [doi: 10.1111/dme.14534] [Medline: 33524209]
2. Tang TS, Klein G, Görges M, et al. Evaluating a mental health support mobile app for adults with type 1 diabetes living in rural and remote communities: the REACHOUT pilot study. Diabet Med 2025 Mar;42(3):e15451. [doi: 10.1111/dme.15451] [Medline: 39538423]
3. Chin-Jung L, Hsiao-Yean C, Yeu-Hui C, Kuan-Chia L, Hui-Chuan H. Effects of mobile health interventions on improving glycemic stability and quality of life in patients with type 1 diabetes: a meta-analysis. Res Nurs Health 2021 Feb;44(1):187-200. [doi: 10.1002/nur.22094] [Medline: 33368403]
4. Wang X, Shu W, Du J, et al. Mobile health in the management of type 1 diabetes: a systematic review and meta-analysis. BMC Endocr Disord 2019 Feb 13;19(1):21. [doi: 10.1186/s12902-019-0347-6] [Medline: 30760280]
5. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297. [doi: 10.2196/12297] [Medline: 30664494]
6. Sun C, Malcolm JC, Wong B, Shorr R, Doyle MA. Improving glycemic control in adults and children with type 1 diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes 2019 Feb;43(1):51-58. [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
7. Pi L, Shi X, Wang Z, Zhou Z. Effect of smartphone apps on glycemic control in young patients with type 1 diabetes: a meta-analysis. Front Public Health 2023;11:1074946. [doi: 10.3389/fpubh.2023.1074946] [Medline: 37064701]
8. Adu MD, Malabu UH, Malau-Aduli AEO, Malau-Aduli BS. The development of My Care Hub mobile-phone app to support self-management in Australians with type 1 diabetes or type 2 diabetes. Sci Rep 2020 Jan 8;10(1):7. [doi: 10.1038/s41598-019-56411-0] [Medline: 31913296]
9. Castensøe-Seidenfaden P, Reventlov Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124. [doi: 10.2196/mhealth.8137] [Medline: 29061552]
10. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12(12):CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
11. Bellei EA, Biduski D, Lisboa HRK, De Marchi ACB. Development and assessment of a mobile health application for monitoring the linkage among treatment factors of type 1 diabetes mellitus. Telemed J E Health 2020 Feb;26(2):205-217. [doi: 10.1089/tmj.2018.0329] [Medline: 30724717]
12. Tang TS, Yip AKW, Klein G, et al. Training peers to deliver mental health support to adults with type 1 diabetes using the REACHOUT mobile app. Diabet Med 2024 Mar;41(3):e15210. [doi: 10.1111/dme.15210] [Medline: 37634222]
13. Tang TS, Seddigh S, Halbe E, Vesco AT. Testing 3 digital health platforms to improve mental health outcomes in adults with type 1 diabetes: a pilot trial. Can J Diabetes 2024 Feb;48(1):18-25. [doi: 10.1016/j.jcjd.2023.08.006] [Medline: 37625504]
14. Muijs LT, de Wit M, Knoop H, Snoek FJ. Feasibility and user experience of the unguided web-based self-help app "MyDiaMate" aimed to prevent and reduce psychological distress and fatigue in adults with diabetes. Internet Interv 2021 Sep;25:100414. [doi: 10.1016/j.invent.2021.100414] [Medline: 34401373]
15. Rondags SMPA, de Wit M, Twisk JW, Snoek FJ. Effectiveness of HypoAware, a brief partly web-based psychoeducational intervention for adults with type 1 and insulin-treated type 2 diabetes and problematic hypoglycemia: a cluster randomized controlled trial. Diabetes Care 2016 Dec;39(12):2190-2196. [doi: 10.2337/dc16-1614] [Medline: 27797931]

XSL•FO
RenderX

16. Yakubu TI, Pawer S, West NC, Tang TS, Görges M. Impact of digitally enabled peer support interventions on diabetes distress and depressive symptoms in people living with type 1 diabetes: a systematic review. Curr Diab Rep 2025 Dec;25(1):1. [doi: 10.1007/s11892-024-01560-4]
17. Tamhane S, Rodriguez-Gutierrez R, Hargraves I, Montori VM. Shared decision-making in diabetes care. Curr Diab Rep 2015 Dec;15(12):112. [doi: 10.1007/s11892-015-0688-0] [Medline: 26458383]
18. Bradway M, Morris RL, Giordanengo A, Årsand E. How mHealth can facilitate collaboration in diabetes care: qualitative analysis of co-design workshops. BMC Health Serv Res 2020 Nov 30;20(1):1104. [doi: 10.1186/s12913-020-05955-3] [Medline: 33256732]
19. Moore A, Cepeda J, Franklin B, Abreu G, Dorth S, Barkley E. Dominican adolescents' preferences for content, design, and functionality of a mobile application for type 1 diabetes mellitus self-management. Can J Diabetes 2023 Dec;47(8):665-671. [doi: 10.1016/j.jcjd.2023.07.006] [Medline: 37481124]
20. Wilson LM, Tyler N, Jacobs PG, et al. Patient input for design of a decision support smartphone application for type 1 diabetes. J Diabetes Sci Technol 2020 Nov;14(6):1081-1087. [doi: 10.1177/1932296819870231] [Medline: 31441336]
21. Holtz BE, Mitchell K. Best practices for developing a diabetes mHealth app. J Diabetes Sci Technol 2024 Jan;18(1):39-45. [doi: 10.1177/19322968231195849] [Medline: 37644831]
22. Alfonsi JE, Choi EEY, Arshad T, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e22074. [doi: 10.2196/22074] [Medline: 33112249]
23. Chatzakis C, Floros D, Papagianni M, et al. The beneficial effect of the mobile application Euglyca in children and adolescents with type 1 diabetes mellitus: a randomized controlled trial. Diabetes Technol Ther 2019 Nov;21(11):627-634. [doi: 10.1089/dia.2019.0170] [Medline: 31335204]
24. Blanchard J, Ahmed S, Clark B, Sanchez Cotto L, Rangasamy S, Thompson B. Design and testing of a smartphone application for real-time tracking of CSII and CGM site rotation compliance in patients with type 1 diabetes. J Diabetes Sci Technol 2024 Jul;18(4):937-945. [doi: 10.1177/19322968221145178] [Medline: 36539997]
25. Shetty VB, Fried L, Roby HC, et al. Development of a novel mobile health app to empower young people with type 1 diabetes to exercise safely: co-design approach. JMIR Diabetes 2024 Jul 30;9:e51491. [doi: 10.2196/51491] [Medline: 39078700]
26. Ng JYY, Ntoumanis N, Thøgersen-Ntoumani C, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
27. Slack. URL: https://app.slack.com/plans/T078CK2H98S [accessed 2024-11-05]
28. Rocket.Chat pricing. Rocket.Chat. URL: https://www.rocket.chat/pricing [accessed 2024-11-05]
29. Mattermost pricing. Mattermost. URL: https://mattermost.com/pricing/ [accessed 2024-11-05]
30. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mHealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019;24(4):254-269. [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
31. Lam D, Jhajj P, Sherifali D, Chen FS, Tang TS. Exploring the REACHOUT mental health support app for type 1 diabetes from the perspectives of recipients and providers of peer support: qualitative study. JMIR Diabetes 2026 Jan 21;11:e72779. [doi: 10.2196/72779] [Medline: 41564292]
32. Clegg D, Barker R. Case Method Fast-Track: A Rad Approach: Addison-Wesley Publishing Company; 1994. URL: https://books.google.co.in/books/about/CASE_Method_Fast_track.html?id=86ZfQgAACAAJ&redir_esc=y [accessed 2026-02-21]
33. Yakubu TI, Jhajj P, Pawer S, et al. Exploring the needs and preferences of users and parents to design a mobile app to deliver mental health peer support to adolescents with type 1 diabetes: qualitative study. JMIR Diabetes 2025 Jan 20;10:e64267. [doi: 10.2196/64267] [Medline: 39832366]
34. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. Presented at: Python in Science Conference; Aug 19-24, 2008. [doi: 10.25080/TCWV9851]
35. Morita PP, Cafazzo JA. Challenges and paradoxes of human factors in health technology design. JMIR Hum Factors 2016 Mar 1;3(1):e11. [doi: 10.2196/humanfactors.4653] [Medline: 27025862]
36. Brooke J. SUS: a "quick and dirty" usability scale. In: Usability Evaluation in Industry, 1st edition: CRC Press; 1996, Vol. 189:4-7 URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-02-21]
37. Chin JP, Diehl VA, Norman LK. Development of an instrument measuring user satisfaction of the human-computer interface. 1988 Presented at: The SIGCHI Conference; May 1-5, 1988. [doi: 10.1145/57167.57203]
38. Sauro J, Lewis JR. Standardized usability questionnaires. In: Quantifying the User Experience: Practical Statistics for User Research: Morgan Kaufmann; 2012:185-240 URL: https://www.sciencedirect.com/book/monograph/9780123849687/quantifying-the-user-experience?utm_source=chatgpt.com [accessed 2026-02-21] [doi: 10.1016/B978-0-12-384968-7.00008-4]

Abbreviations

AWS: Amazon Web Services

XSL•FO
RenderX

mHealth: mobile health
RCT: randomized controlled trial
T1D: type 1 diabetes

Edited by R Correa, N Cahill; submitted 24.Jan.2025; peer-reviewed by C br, C Pankhurst, T David, U Hemetek; accepted 06.Feb.2026; published 20.Mar.2026.

Please cite as:
Lee S, Sidhu B, Johal P, Azhar A, Sujan J, Görges M, Tang TS
T1D REACHOUT–A Mobile App to Deliver Peer-Led Mental Health Support to Adults Living With Type 1 Diabetes: Co-Design and Development Process
JMIR Diabetes 2026;11:e71733
URL: https://diabetes.jmir.org/2026/1/e71733
doi:10.2196/71733

© Sadie Lee, Baray Sidhu, Parteek Johal, Ayman Azhar, Jonath Sujan, Matthias Görges, Tricia S Tang. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Experiences of Adults With Type 1 Diabetes Using Digital Health Technology for Diabetes Self-Care: Qualitative Study

Divya Anna Stephen[1], PhD; Jan Nilsson[1,2], PhD; Unn-Britt Johansson[3], PhD; Anna Nordin[1], PhD

[1]Department of Health Science, Faculty for Health, Nature and Technology, Karlstad University, Universitetsgatan 2, Karlstad, Sweden

[2]Section for Advanced Nursing, Faculty of Social and Health Science, University of Inland Norway, Elverum, Norway

[3]Department of Health Promoting Science, Sophiahemmet University, Stockholm, Sweden

Corresponding Author:
Anna Nordin, PhD
Department of Health Science, Faculty for Health, Nature and Technology, Karlstad University, Universitetsgatan 2, Karlstad, Sweden

Abstract

Background: Type 1 diabetes is a constraining disease due to the burden of its management, and diabetes outcome largely depends on the effectiveness of diabetes self-care. Digital health technology (DHT), which includes continuous glucose monitoring, insulin delivery devices, and related mobile health apps, can support diabetes self-care and thereby improve diabetes outcomes. In literature, experiences with the use of DHT vary widely among people with diabetes and are a less studied area among adults with type 1 diabetes.

Objective: The study aimed to explore experiences of using DHT for diabetes self-care among adults with type 1 diabetes.

Methods: A qualitative design with an inductive approach was used. Adults with type 1 diabetes who are users of DHT and could understand Swedish were included in the study. Participants were recruited primarily via digital advertisements through social media. A convenient sampling method was used. Data were collected through open-ended questions in a web-based survey (autumn 2022) and 2 digital group interviews (autumn 2024). The survey questionnaire and interview guide attempted to capture positive and negative experiences of using DHTs for diabetes self-care through personally relevant incidents and behavioral details. Data from a total of 161 participants (n=156 survey participants and n=5 interview participants), using 1 or more forms of DHTs, were included in the study. Data were analyzed using qualitative content analysis with an inductive approach as per Graneheim and Lundman. The data in this study generated 324 meaning units relevant to the aim.

Results: The participants experienced using DHTs in diabetes self-care as a balancing act between feeling empowered and feeling exasperated. This is described under 5 categories: promoting autonomy in daily life, self-awareness through collaborative learning, feeling secure, tackling technical challenges and the need for support, and navigating the burden of psychosocial challenges. DHTs were experienced as empowering when they supported autonomy in daily life, enhanced self-awareness through collaborative learning, and fostered a sense of security. However, having to tackle technical challenges and the need for support, and navigating the burden of psychosocial challenges, led to feelings of exasperation. The exasperating experiences hindered participants from experiencing a full sense of empowerment with DHT use.

Conclusions: This study sheds light on both positive and negative experiences of using DHTs for diabetes self-care in a real-life setting. The exasperating experiences may widen the digital health inequities and therefore are important to address. Improving technological literacy and ongoing support from health care or device manufacturers may help users to address exasperating experiences. Further studies are needed to validate our findings.

(JMIR Diabetes 2026;11:e79704) doi:10.2196/79704

KEYWORDS

digital technology; experiences; qualitative research; self care; type 1 diabetes mellitus

Introduction

Type 1 diabetes places a unique burden of management on the individual. It is an autoimmune disorder leading to severe endogenous insulin deficiency [1]. It requires the individual to be always in control of the disease. The latter can be experienced as constraining life [2]. The incidence and prevalence of type 1 diabetes has increased globally [3] with Sweden having one of the highest age-standardized incidences both among children [4] and adults [5]. Effective diabetes self-care plays a role in determining the outcome of diabetes treatment [1]. The 7 recommended diabetes self-care behaviors are healthy coping, healthy eating, being active, taking medication, monitoring, reducing risk, and problem solving [6]. These self-care behaviors and the interplay between them affect glycemic control and

XSL•FO
RenderX

occurrence of complications [7]. However, despite the relentless self-care measures, people still experience loss of control due to their inability to address unforeseen factors that impact glucose levels [2]. New approaches and devices should be incorporated into diabetes care to minimize the burden of living with type 1 diabetes [1]. Newer technologies help to synthesize information into an easily understandable and interpretable form, enabling its application to self-care. They can engage, encourage, and motivate people with diabetes to manage their self-care [6].

Digital health technology (DHT)—or diabetes technology—includes continuous glucose monitoring (CGM), insulin delivery devices, and related mobile health (mHealth) apps. Insulin delivery devices include continuous subcutaneous insulin infusion (CSII), automated insulin delivery (AID) systems, and smart or connected insulin pens [8,9]. The DHTs have been found to improve glucose outcomes in people with diabetes [8,10]. In addition, the data generated by them is perceived to make life easy for people with type 1 diabetes [11]. Various forms of insulin delivery devices and CGMs have been available to people with type 1 diabetes from the late 1970s [12] and 1999 [13] respectively. However, after the 2010s, the technological developments have continuously brought new features and functionalities [12,13]. Although this development potentially makes diabetes self-care easier, it may also act as a barrier to effective uptake [8]. Some commonly reported barriers and burdens to DHT use were cost, numerous alarms, discomfort, need for engagement, and workload [14].

Experiences offer important information about the nature of people's needs pertaining to various interventions [15], including DHTs. DHTs such as CGMs [16,17], CSII [18], and AIDs [19,20] were experienced as tools to help reduce glucose variability, improve glucose control [17,19,20], and overall diabetes burden [20], offering people with diabetes more control over their own lives [16-18]. The DHTs were experienced as convenient, enhancing self-care engagement, and providing confidence [16,17]. However, DHTs can place a significant burden on people due to the need for continuous interaction with devices when it comes to placement, calibration, and troubleshooting [19,21], and data overload [18,21]. They also cause challenges for people by drawing unwanted attention [18,21], disruptive alarms and alerts causing disturbance to others as well as to daily life [16,19-21], body image issues, and issues with devices during sports [21]. Other challenges were physical discomfort [16,18] due to their invasive nature, and skin reactions [21]. Additionally, unfamiliarity with technology, malfunctioning and inaccuracy of devices, mistrust in algorithms, economic constraints, supply deficit [16], sensor problems, not matching people's expectations [19], and uneasiness in relinquishing control to the system [19,20] were also reported challenges. The challenges with DHTs can be experienced as a burden [21]. Thus, experience with the use of DHT varies widely among people with diabetes.

DHTs have evolved rapidly, which has had a positive impact on diabetes self-care and glucose outcomes. However, people's expectations, the current state of technology, as well as access to technical support and health care support may differ in a real-world setting compared to randomized controlled trials. Therefore, it is essential to explore people's experiences of using DHTs in day-to-day life. This knowledge may contribute toward understanding and improving the gaps that prevent effective use of DHTs in diabetes self-care. Nevertheless, studies in this area in a long-term, real-world setting are rare. In Sweden, the DHTs are available to people with type 1 diabetes through a publicly funded high-cost protection scheme [22,23] and therefore usage rates are high [24,25]. This provides a unique opportunity to explore experiences among long-term users. Thus, the aim of this study was to explore experiences of using DHT for diabetes self-care among adults with type 1 diabetes.

Methods

Study Design

The study had a qualitative design with an inductive approach and was reported as per consolidated criteria for reporting qualitative research (COREQ) guidelines [26]. It includes data from 2 open-ended questions from a web-based survey and 2 group interviews to describe experiences of using DHT for type 1 diabetes self-care.

Participants

Adults (18 years or older) with type 1 diabetes who were users of DHT and could understand Swedish were included in the study. Participants were chosen by convenience sampling method. A total of 162 participants answered the open-ended questions in the survey. However, data from 6 (3.7%) participants were excluded, as they did not describe an experience. This led to open-ended question data from 156 (96.3%) participants being included in the study. For the group interview, 12 participants had registered interest and were invited to participate. However, 7 (58.3%) participants could not participate for personal and technical reasons, resulting in 5 (41.7%) interview participants. Thus, a total of 161 participants provided data to this study. Included participants used 1 or more forms of DHT. Brands of DHT and mHealth app features used by participants are available in Table S1 and Table S2 in Multimedia Appendix 1.

Data Collection

Survey

Participants were recruited between September and November 2022, primarily through digital advertisements on social media and completed a web-based survey (Survey&Report by Artisans media). Additionally, study advertisements were done at associations for people with diabetes in Sweden and a regional hospital. The survey included screening questions on age, diabetes type, and pregnancy status to determine participants' eligibility. The survey automatically closed if any of the exclusion criteria (age <18 years, diabetes type 2, or pregnancy) were met. More information on the survey recruitment strategy is available in a previously published paper [25]. Two open-ended questions in a web-based survey, which was part of a larger study [25,27], provided data to this paper. Although paper surveys were available, their uptake was minimal, and none of the respondents provided answers to the open-ended questions.

XSL•FO
RenderX

The phrasing of the interview questions was inspired by the critical incident method [28,29]. The questionnaire began by asking participants to describe an experience where they were *satisfied* with their use of the mHealth app for diabetes self-care, and then to describe an experience where they were *dissatisfied*. They were then asked to relate details about the particular incident or situation and describe how they managed it, how it ended, and what they learned. Phrasing the questions this way helped participants recall and describe their positive and negative experiences using DHTs for diabetes self-care through personally relevant incidents and behavioral details [30]. The questions were face-validated among 9 adults with type 1 diabetes and 4 diabetes nurses, where after minor changes were made.

Group Interviews

An initial review of the web-based survey data was conducted by the first (DAS) and last (AN) authors to assess the need for further enrichment and updates, which was deemed necessary. Thus, 2 group interviews were conducted between August and September 2024. The participants for the group interview were recruited via digital advertisements through Facebook and Instagram. The interested participants could leave their informed consent and contact details through a digital link (Survey&Report by Artisans media) in the advertisement. The group interviews included 2 and 3 participants each. The participants were sent the interview invitation through email. The video conferencing tool Zoom Workplace was used to conduct the interviews. The end-to-end encryption feature in Zoom was activated to secure the room and ensure data privacy.

A semistructured interview guide, developed based on the open-ended questions in the survey study, was used for the interview. The questions were phrased slightly differently based on the data from the survey in order to include digital devices along with mHealth apps. Digital devices were added as survey answers showed that participants could not differentiate the functions of mHealth apps from those of their digital devices. The interview began in a similar way to the survey questionnaire, with an overarching question asking them to describe experiences of using mHealth apps and digital technology to manage type 1 diabetes. Similar to the questionnaire, this was followed by sub-questions. Follow-up questions were asked as needed: Can you tell me more? Can you elaborate on that? AN (PhD in nursing science and associate professor) was the main interviewer and has previous experience of conducting qualitative interviews and knowledge in the field of digital health innovations. DAS, a PhD student, acted as the observer and note taker. Both the interviewers were female, registered nurses and had no previous relationship with the participants. The interviews, which included only the interviewers and participants, lasted between 60 to 90 minutes and were audio recorded using a separate dictaphone. The audio recordings were transcribed verbatim by a transcription agency, which had a data processing agreement with Karlstad University. The interview transcripts were not returned to the participants for correction or comments.

Data Analysis

Qualitative content analysis as per Graneheim and Lundman [31] was used to analyze the data. An inductive or data-driven approach characterized by the search for patterns was used [32]. AN and DAS initially reviewed survey data to determine the necessity for further data collection. However, no coding or in-depth data analysis was done on the survey data until the end of data collection through group interview. DAS and AN manually performed the initial coding of all data in Microsoft Excel using a convergent design, where both survey and interview texts were coded from scratch. Any disagreements were discussed among all 4 authors until consensus was reached. No qualitative data analysis software was used.

The survey text was read repeatedly and answers that described whole or part of an experience were assigned as a meaning unit by DAS. The meaning units were then condensed and coded close to the text by DAS and AN. Similarly, the interview text was read as a whole several times by DAS before extracting meaning units. It was then condensed and coded close to text by DAS and AN. The codes were then grouped based on their similarities and differences into 5 categories and a theme. Data saturation was discussed post hoc and was deemed as achieved when categories were found repeating [33]. A back-and-forth process of going between data, codes, and categories was carried out to ensure codes fit the category. The categories and themes were discussed among all 4 others until consensus was achieved. The data in this study gave rise to 324 meaning units with relevance to the aim, of which one-third were from group interviews [31,32]. The coding and the categorization were discussed several times in the research group and at the research seminars at the research school to strengthen confirmability and credibility. All 3 authors except the first author were experienced in qualitative research and guided the first author through the process. See Table 1 here and Table S3 in Multimedia Appendix 2 for examples of the data analysis process. Participants were not consulted for feedback on the findings.

XSL•FO
RenderX

Table . Example of the data analysis process.[a]

Meaning unit	Condensed meaning unit	Code	Subcategory	Category
I have always had problems with low blood sugar at night...usually because I exercise later in the evening. I am very happy that my device wakes me up at night when my blood sugar is low so that nothing serious happens and next day is not ruined by consequences of my low blood sugar at night. [P283, 34 y, female, using CGM[b]]	Waking me up at night with low blood sugar, preventing complications as well as the next day being subsequently ruined	Low blood sugar alarms at night prevent complications and the next day being ruined	Alarms and warnings prompt timely intervention	Promoting autonomy in daily life
Because it's no fun sitting in a court that's like a library...And then comes that beep and then the phone rings five seconds right after. Not the right context...I solved it by looking terribly embarrassed and then in the break I bolted (laughter). It's like maybe a situation where you want to close your eyes.[P452, 46 y, male, using CGM and CSII[c]]	Alarm from device and later from phone during a session at court led to feeling embarrassed and stressed	Alarms at inconvenient places are embarrassing and stressful	Source of social awkwardness	Navigating the burden of psychosocial challenges

[a]Theme: A balancing act between feeling empowered and feeling exasperated.

[b]CGM: continuous glucose monitoring.

[c]CSII: continuous subcutaneous insulin infusion.

Ethical Considerations

This study follows the ethical guidelines as per the Helsinki declaration [34]. Participation in the study was voluntary, and no compensation was provided. Participants were given oral and written information about the study and their right to withdraw participation at any time. Informed consent was obtained from the participants digitally via the survey tool. The data related to this study were stored on Sunet Drive, a secure cloud storage service, and on an appropriately encrypted USB drive kept in a locked archive at Karlstad University, accessible only to the research team. In accordance with institutional guidelines, the data will be retained at Karlstad University for a period of 10 years. The Swedish ethical review authority approved the study before data collection commenced (Dnr: 2021-05337-01, Dnr: 2022-04079-02, Dnr 2024-02691-02).

Results

In total, 161 participants provided data for this study. The participant characteristics are summarized in Table 2.

XSL•FO
RenderX

Table . Characteristics of participants who answered the open-ended question in the survey and participated in the focus group interviews.

Characteristics	Values	
	Survey (n=156)	Interview (n=5)
Age (y), median (range)	36.5 (18-79)	48 (25 - 68)
Duration of diabetes (y), median (range)	18 (<1 - 72)	5 (1-43)
Duration of digital health device usage (y), median (range)	__[a]	5 (1-8)
Gender, n (%)		
Women	104 (66.7)	2 (40)
Men	52 (33.3)	3 (60)
Education level, n (%)		
University level education	93 (59.6)	—
Primary/secondary school	63 (40.4)	—
Digital devices used[b], n (%)		
Blood glucose monitor	79 (50.6)	5 (100)
Continuous glucose monitoring	144 (92.3)	4 (80)
Continuous subcutaneous insulin infusion	45 (28.8)	1(20)
Automated insulin delivery systems	42 (26.9)	1 (20)
Smart insulin pens	19 (12.2)	—
Mobile health apps	156 (100)	5 (100)

[a]Not available.

[b]Participants could use more than one digital device.

Participants' Experiences of Using DHTs

Theme: A Balancing Act Between Feeling Empowered and Feeling Exasperated

The participants experienced using DHTs in diabetes self-care as a balancing act between feeling empowered and feeling exasperated. This is described under five categories: promoting autonomy in daily life, self-awareness through collaborative learning, feeling secure, tackling technical challenges and need for support, and navigating the burden of psychosocial challenges. DHTs were experienced as empowering as they helped people take control over their lives by promoting autonomy in daily life, self-awareness through collaborative learning, and making them feel secure. However, participants have also described having to tackle technical challenges and the need for support and navigating the burden of psychosocial challenges, which left them feeling exasperated. Thus, participants conveyed an experience of having to balance between feeling empowered and feeling exasperated. The exasperating experiences prevented them from experiencing a full sense of empowerment with DHT use. Textbox 1 depicts the various categories and Table 3 depicts subcategories and count of codes within each category.

Textbox 1. Summary of participants' experiences of using digital health technologies for self-care in type 1 diabetes.

A balancing act between feeling empowered and feeling exasperated includes the following categories:

- Tackling technical challenges and the need for support
- Promoting autonomy in daily life
- Self-awareness through collaborative learning
- Navigating the burden of psychosocial challenges
- Feeling secure

XSL•FO
RenderX

Table . Subcategories and count of codes within each category (n=324).

Categories and subcategories	Count of codes, n
Tackling technical challenges and the need for support	145
Hassles with DHT[a] malfunction and usability	
Glucose value inaccuracies	
Unreliable alarms and warnings	
Health care support need	
Promoting autonomy in daily life	108
Aiding glycemic control	
Alarms and warnings prompt timely intervention	
Eases insulin dosing	
Self-awareness through collaborative learning	28
Understanding factors affecting glucose variability	
Implementing self-care changes based on value assessment	
Eases discussion on diabetes self-care	
Navigating the burden of psychosocial challenges	27
Psychological impact and response	
Source of social awkwardness	
Feeling secure	16
Experiencing stability in their lives	
Safety and control over their disease	

[a]DHT: digital health technology.

Category 1: Tackling Technical Challenges and the Need for Support

A large number of participants with type 1 diabetes described their experiences of tackling technical challenges and need for support. They described hassles with DHT malfunction and usability, glucose value inaccuracies, unreliable alarms and warnings, and health care support need. Tackling technical challenges with DHT and the need for support led to exasperation.

DHT malfunctions experienced by participants ranged from software-related issues such as failure in updating values and display inaccuracies to device dysfunctions. Unreliability in alarms and warning systems was also described, including false alarms and failing to alert due to either DHT malfunctions or design flaws. Alarms were also experienced as developmentally immature when they could not be personalized or overridden. Identified causes of DHT malfunction described by the participants were either related to sensor issues, dysfunctions when approaching end of lifespan of device parts, incompatibility with extreme external temperatures, physiologic changes (eg, infection) or diet causing rapid glucose fluctuations, device connectivity issues with mHealth apps or other connected devices, lack of device compatibility with other devices or newer software, and usability issues (eg, navigational difficulties and limited personalization options).

> *I also noticed that this summer when it was like 30 degrees in the sun and you sat outside, then..for me the values went upwards. It looked as if I had tucked in myself with 1 kg sugar. And then you come in and it takes five minutes and it looks like it is reversing down but it stays stable,...it was really something that you learned that one needs to adjust for both heat and cold...it should be just moderate all the time.* [P455, 25 y, male, using CGM]

Participants experienced disruptions in connected device functions (eg, AID system), lack of access to glucose values, and need for reliance on nonapproved apps due to technical challenges. In addition, participants reported technical challenges as contributing to a cascade of poor self-care decisions—beginning with uncertainty around glucose values and insulin dosage estimation, which led to incorrect dosing, potentially resulting in compensable or noncompensable hyperglycemia or hypoglycemia, and in some cases, the need for emergency care. Participants described eventually learning to navigate the technical challenges by using DHTs as a guide and relying on their own body symptoms and sensations for accurate feedback. When a mismatch in DHT information and bodily symptoms was noticed by participants, it was counterchecked with glucose graphs or with blood glucose monitors. Some malfunctions required the participants to correct underlying causes such as relieving pressure on devices and change of sensor or parts of insulin delivery devices. The latter, however, was reported by the participants as requiring backup material, being uncomfortable if in public eye, irritating to the skin, and required waiting out the “warm up” period. To keep

XSL•FO
RenderX

using the DHTs, participants narrated the need for relentlessly being in control through careful organization and preparation. Even then, the DHTs could still fail, causing disruptions to daily routines. DHTs were described as “a shackle” when preventing spontaneous activities due to a lot of backup need.

> *But it’s a big difference in some way, I feel a little more shackled and a little more exposed when I'm dependent on the [insulin] pump...with the pump there is so much, there are infusion sets, there are needles, there is the insulin, it has to be cold and...you have to charge it...There it can very easily tip over and become stressful instead.* [P452, 46 y, male, using CGM and CSII]

Participants felt that these technical challenges induced value speculations, counterchecks, wrong self-care interventions, and subsequent correctional measures that were unnecessary and inconvenient, disrupting routines, stressful, tiresome, and undermining trust in DHTs. The absence of prompt technical assistance was described by participants as a factor contributing to disengagement and eventual discontinuation of the DHT. Some participants shared that they received minimal support from healthcare for DHT use, often turning to social media for help. Technical challenges frequently occurred outside of health care providers’ consultation hours, highlighting a lack of around-the-clock support. Using DHTs with minimal health care support was compared by participants to that of a “technical sport—a sport requiring knowledge of technology, extensive exploration and adaptation without which you lag in the race.” Participants expressed a need for support in better understanding the technology and selecting the most suitable option from those available. Participants felt that adequate support would prevent them from feeling chained to ill-fitting DHTs. Some found it helpful to have consistent contact with the health care personnel who recommended the DHTs. However, participants voiced that this level of support requires health care professionals to maintain up-to-date knowledge of available DHTs. Participants also described a lack of knowledge among health care personnel in primary care and home health care when it came to caring for people with type 1 diabetes using DHTs.

> *And the thing is also that nothing with technology or with diabetes happens when the diabetes nurse has opening hours between 8 and 9 in the morning...but it usually occurs in the evenings...you sometimes forget a little human and as I said, it takes quite a lot and I feel...if you do not really have that interest [yourself]...you easily...fall behind in it.* [P454, 48 y, female, using CGM]

Category 2: Promoting Autonomy in Daily Life

Participants with type 1 diabetes felt that DHTs promoted autonomy in daily life by aiding glycemic control, easing insulin dosing, and prompting timely intervention through alarms and warnings. Participants described this as being achieved through frequent glucose checks and tracking, decision support to intervene in time through alerts, alarms, and warnings, and assisting with insulin administration in everyday life as well as during special situations (eg, infection or while having a multicourse meal). In comparison to a regular finger prick test with blood glucose monitor, participants felt that DHTs promoted autonomy by making self-care monitoring easier, frequent, discreet, pain-free, and always accessible. Through graphs, trends, and alarms, participants described receiving a comprehensive and uninterrupted overview of glucose values or insulin dose, which was unattainable through occasional finger prick tests or glycated hemoglobin (HbA_{1c}) tests. This comprehensive and uninterrupted diabetes monitoring and management was identified by participants as helpful to keep glucose values in normal range and improve “time in range.” Participants mentioned this as reducing worry of hypo- or hyperglycemia while being engaged in various activities or work and at night. Therefore, DHTs were experienced as a lifeline working around the clock, preventing complications or disruptions to daily routines. By giving enough time to intervene to correct glucose levels, they provided participants with the freedom of activity. Participants described that DHTs enabled them to live their lives without having to think too much about their disease.

> *I think Pump...it’s really great, both accelerates and brakes, it helps to go down at night...then it stops the insulin needed...[when] you're almost low so it...starts warning with four...it manages the night...but otherwise...I would have been woken up in the night...or something else could have happened. So that’s the advantage.* [P453, 68 y, male, using AID]

> *Every time it alerts for the blood sugar...high value over 11 mmol...I have time to correct before it gets higher. Thanks to this...HbA1c went from 70 to 54 in about 1 year.* [P204, 26 y, female, using CGM and CSII]

Category 3: Self-Awareness Through Collaborative Learning

Participants with type 1 diabetes voiced that DHTs contributed to a greater self-awareness about their disease and the self-care strategies required. Achievement of this increased self-awareness was described by participants as being facilitated by participants’ collaborative engagement and learning with DHTs. The participants described this experience under the subcategories: understanding factors affecting glucose variability, implementing self-care changes based on value assessment, and easing discussion on diabetes self-care with health care personnel.

DHTs were experienced as helping participants learn about diabetes and the impact on glucose levels of various factors such as exercise, insulin dosing, stress, menstrual cycle, and food, thereby facilitating self-care. Participants described achieving this learning through the DHTs’ graphical display and trends, which help find patterns, recurring events, habits, and reason for deviant values. The insight gained helped participants make self-care changes and prevent glucose variability. Alarms and predictions helped them learn differences in body symptoms with deviant glucose values. In contrast to periodic finger prick testing using blood glucose monitors, DHTs helped users learn to time the effects of insulin and food. By aiding participants with analyzing and evaluating the interplay between various factors influencing glucose levels,

XSL•FO
RenderX

DHTs helped lower HbA_{1c}. Participants also felt that the uploaded or shared data from the DHTs made it easier to discuss diabetes self-care with health care personnel even without in-person visits. Graphs from devices were reported to act as a discussion tool in health care meetings.

> *By continuously reviewing graphs, I was able to see patterns in my daily curve and deduce my high blood sugar readings to previously low readings due to low food intake/lack of snacks. Through this, I was able to make adequate adjustments to reduce the occurrence of both low and high blood sugar values and ensure a more stable curve.* [P388, 30 y, female, using AID]

> *I exercise a lot...demanding work and comes up with spontaneous things. With CGM and connection to watch and alarm, I have enough knowledge for good planning...I have above all learned how to plan exercise and insulin / food before and during sessions that can last many hours. Then how these affect my long-acting insulin, which must be lowered by up to 50% after intensive exercise. Decent control of the curves...along with the motivation to exercise.* [P63, 48 y, male, using CGM]

Category 4: Navigating the Burden of Psychosocial Challenges

Experiences of psychosocial challenges while using DHTs were also described by participants with type 1 diabetes, under subcategories psychological impact and response to DHTs and a source of social awkwardness. Participants described feeling compelled to maintain values within the target range with DHT use, making them fixated on achieving the perfect “time in range” values and overly trusting DHTs above bodily symptoms. Participants felt that this led to stress and a sense of self-competition, particularly during the initial phase of technology use. Over time, some participants learned to ease this pressure and place less importance on staying within the perfect range. Such trust in DHTs above bodily symptoms caused unnecessary anxiety in participants about DHTs failing to work or when seeing false “in range” percentages. The obsession to remain in target range and high focus on DHTs was also described by participants as leading to overcorrection, unnecessary interventions, incorrect insulin dosing, fluctuating values, and unwanted consequences. Additionally, it could lead to values, alarms, and reminders taking over every part of participants’ life. DHTs were thus experienced as disturbing peace and quiet, causing enormous stress and frustration, which eventually led to technology fatigue. Taking a day off DHTs helped some participants realize that they are no more than aids to diabetes self-care.

> *Sometimes I can get very obsessive and frustrated when I'm high and the curve doesn't want to turn as quickly as I expected, I can be tempted to take more doses of the insulin. When the curve then turns, it goes too fast and I have to correct with extra carbohydrates.* [P252, 56 y, female, using CGM]

> *It gives a huge stress to follow glucose levels in the app all the time along with the warning/alert sounds day and night when blood sugar sways. One does not get any peace and quiet. This gives an enormous stress along with other things one should achieve/tread through in life.* [P16, 49 y, female, using CGM and smart insulin pen]

DHT can also cause social hassles and awkwardness. Participants described frustration over uneducated comments from others about DHTs as well as 1-sided (ie, only the positives) demonstrations of DHTs on social media. Another problem described was that different counties offer different DHTs, making it difficult to move to a different county after choosing a certain DHT. Placement and access of DHT while in public places or events were also experienced as challenging due to physical exposure. One participant felt undignified when people showed aversion and judgment toward DHT alarms. Lack of customization like snoozing, overriding, or volume adjustments for alarms was described by participants as causing embarrassment, disturbance in public places, and in extreme cases, discontinuation of DHT.

Category 5: Feeling Secure

Participants with type 1 diabetes experienced feeling secure when using DHT for diabetes self-care. They described this under experiencing stability in their lives and safety and control over their disease with DHT use. Participants reported a sense of safety since alarms would alert them to deviant glucose values in work-related meetings and at night when asleep. This, in turn, gave them a feeling of being in control, stability, and reduced stress about complications such as diabetes-related coma. Some participants described how the follower function saved their lives by alerting their relatives to intervene in the event of low blood sugar, which went unnoticed by the participants.

> *The follower function where my daughters helped me with too low blood sugar to avoid going into a [diabetes-related] coma! I did not alert myself...about the low blood sugar, or I became low too fast and could not react. I...don't feel sensations anymore, until the blood sugar is too low.* [P382, 79 y, male, using CGM and smart insulin pen]

> *Alarms about low glucose values at night have helped to reduce stress about dangerous situations occurring when I am not aware.* [P305, 22 y, male, using CGM]

Discussion

Principal Results

This study examines adults’ experiences of using DHT for diabetes self-care in real-life settings. To our knowledge, research in this area among adults with type 1 diabetes remains limited. The findings reveal that these experiences reflect a balancing act between feeling empowered and feeling exasperated, discussed under the following categories: promoting autonomy in daily life, self-awareness through collaborative learning, feeling secure, tackling technical challenges and need for support, and navigating the burden of psychosocial challenges. In line with our findings, use of DHTs

XSL•FO
RenderX

[35], CGM [36], and AID [37] has been reported as empowering people with type 1 diabetes [36,37]. However, this study's findings also indicate that participants were not able to fully feel empowered due to feelings of exasperation with DHTs. Another related study also found no association between the total diabetes empowerment score and the DHT factors among the same sample [25]. Therefore, larger studies exploring empowerment using well-established patient-reported outcome measures among this population are needed to validate our findings.

In this study, DHTs were experienced as promoting autonomy in daily life. Other studies have also reported findings in relation to this category, such as feeling liberated [17] and getting part of one's life back [17], aiding glucose control [16-18] and reducing disease-related restriction on activities [38]. Participants experienced feeling secure in their daily lives and with their diabetes self-care when using DHTs, which is in line with findings from other studies among adults with type 1 diabetes [2,17]. Contrary to our results, 1 study reported participants' reluctance to use CSII pumps due to a preference for the well-known, technology skepticism, and a belief that it would diminish their sense of security and control [39]. The follower function, in line with our findings, can foster a sense of safety [16] or can be a source of conflict when perceived as surveillance [36]. In our study, DHTs were experienced as increasing awareness of diabetes self-care through collaborative learning, which is similar to findings by Cleal et al [18].

Tackling technical challenges and the need for support is one of the largest categories in this study. Our findings are similar to findings from other studies on experiences of technical challenges like disturbing alarms [16,40], malfunctioning [16,40,41], and value inaccuracy [16]. Descriptions of some incidents falling in this category were challenges, which were known and clearly described by DHT manufacturers in their user manuals. Furthermore, navigating the burden of psychosocial challenges with DHTs is another finding consistent with reports from previous research [21]. Participants' experiences within this category correspond to well-documented constructs like diabetes distress with DHT use [42], alarm fatigue [43], frustration with technology [21,44], and alarm embarrassment [44]. These findings indicate a need for structured periodic education and training on DHTs, as a way to help people with type 1 diabetes effectively tackle technical, emotional, and social challenges related to DHTs [8,21]. This also indicates the need for health care personnel or systems to address this educational need [8]. A continued education program trial for insulin pump users was found to improve the experimental groups' glucose control [45]. Thus, improving people's technological literacy is a way to address disparities in DHT usage [46]. Need for support was a finding in this study, which to our best knowledge was not reported elsewhere. Getting continued support for DHT use is essential in keeping up the morale of people with type 1 diabetes. More studies on the need for support while using DHTs are therefore warranted to validate these findings.

Methodology Discussion and Limitations

Inference from participant data suggests that the study successfully recruited individuals from various regions of Sweden. This enabled the capture of some regional variation in the provision of DHT for individuals with type 1 diabetes [24]. However, collecting explicit information on participants' regions would have ensured greater representativeness. The inclusion of participants of both genders, varying ages, diabetes durations, and educational backgrounds further enhances the transferability of the findings. In addition, the interview participants also varied in their duration of DHT use. However, data on the duration of DHT use were not collected in the survey, which could have further added to the information on transferability. The open-ended question data was pseudonymously collected by the survey tool. Data from interviews was pseudonymized using handwritten codes to ensure that no participant could be individually identified through any of their attributes in the published results. The interview participants were not connected to anyone in the research group. The 4 researchers involved in this study had differing clinical and qualitative experiences. The first and third authors have experience working with patients with type 1 diabetes. Previous studies done by the research group in this area give them a preunderstanding of how people with type 1 diabetes experience using DHTs. This preunderstanding can create a bias during data analysis, but this was minimized by the varied experiences of the other 2 authors, as well as discussions about data analysis in the research group and research seminars at the research school. The preunderstanding was also an advantage when it came to understanding the varied DHT terminology participants used during data collection.

The data for this study were obtained through open-ended questions in a digital survey conducted in autumn 2022 and 2 digital group interviews in autumn 2024. Initially, we aimed to conduct focus group interviews; however, due to the low number of participants, the level of interaction required to generate meaningful data in focus groups could not be ensured [47]. The survey contributed a large number of meaning units, which was a strength of this data source. Nevertheless, these meaning units varied considerably in length (10 - 150 words) and richness. The interview guide, inspired by the critical incident technique, facilitated the capture of experiences from a sufficient number of participants. The inclusion of both positive and negative subquestions ensured that experiences were not described from a single perspective. Several survey responses lacked the outcome component (ie, self-care measures taken to manage the incident) required for inclusion as a critical incident, which is considered essential for credibility in studies using the critical incident technique [29,48]. Consequently, following the critical incident method throughout this study was not considered appropriate. The data were analyzed inductively; however, a deductive approach using a rigorous theoretical framework could have further enriched the findings.

By 2024, newer devices incorporating advanced algorithms had become available through the Swedish health care high-cost protection scheme, compared to 2022. The interviews carried out in autumn 2024 sought to explore potential changes in experiences with digital health technologies during this period.

XSL•FO
RenderX

Nevertheless, given that only 5 participants were included, temporal variations in devices, algorithms, or reimbursement practices may not have been comprehensively captured. An added information on brands and models of each DHT could have provided more insights into participant experiences and enriched the study. Some data on digital technology brands, drawn from participant quotes, is available in Table S1 in Multimedia Appendix 1. However, this list of brands is not comprehensive, as it was not actively collected in the survey. Nevertheless, DHT will continue to evolve, while the categories derived from patient experiences are likely to remain relevant regardless of these changes. As technology advances, there may be variations in the intensity of patients' experiences related to some of the categories. Therefore, the findings remain important despite the absence of detailed brand information on DHTs used by the participants.

Financial constraints affecting DHT uptake have been well documented in literature [16,49,50] and can contribute to diabetes distress, in turn affecting experiences of DHT use [21]. This aspect is not captured in this study results. This could be due to the high cost protection coverage extended to the DHTs by the Swedish public health care system [23]. This financial protection could have played a role in the high uptake of DHTs and shaping participants' experiences reported in this study. Therefore, these results should be generalized with caution to settings with different reimbursement and support infrastructures. Participants were primarily recruited through social media, and data was collected digitally via survey and interviews. This approach enabled the inclusion of participants with diverse demographic backgrounds. However, the recruitment strategy introduces a risk of selection bias toward individuals who are digitally enthusiastic, potentially leading to the underrepresentation of nonusers, lapsed users, and older adults with limited digital literacy. Furthermore, individuals experiencing severe alarm fatigue who discontinued use, as well as those outside Sweden's high-coverage digital context, are less visible in our data. Consequently, the findings of this study primarily reflect the perspectives of active and digitally engaged participants, and caution should be exercised when generalizing to populations with lower digital engagement or different infrastructural conditions. The self-reported diagnosis of type 1 diabetes through an online survey may be considered as another limitation of this study. However, 85% (133/156) of survey participants provided responses that included detailed descriptions of the diabetes technologies they use and how these influence their blood glucose or daily routines. In addition, 135 (86%) out of 156 survey participants reported being diagnosed before the age of 35 years, and 81 (52%) out of 156 before the age of 15 years, which together indicate a high likelihood that the self-reported diagnoses of type 1 diabetes are credible.

Conclusions

The participants experienced the use of DHTs in diabetes self-care as a balancing act between feeling empowered and feeling exasperated. This study sheds light on both positive and negative experiences of using DHTs for diabetes self-care in a real-life setting. The exasperating experiences may widen the digital health inequities and are therefore important to address. Improving technological literacy and ongoing support from health care or device manufacturers may help users tackle technical challenges and navigate the burden of emotional and social hassles. Further research in this area—both within Sweden and internationally—is required to validate these findings and to facilitate their translation into clearly defined user needs and system-level interventions.

Acknowledgments

The authors would like to thank all participants in this study as well as the participants who validated the survey content, the managers and diabetes nurses at the regional diabetes center, and Ellinor Larsson, SUBAB, for help with the language revision of this paper. The generative artificial intelligence technology, Microsoft 365 Copilot, has been used in the refinement and language editing of the manuscript to improve clarity of language. Other than the aforementioned, the authors attest that there was no use of artificial intelligence technology in the research process or generation of text, figures, or other informational content of this manuscript.

Funding

Karlstad University Library provided the open access publication funding. This study was funded by Karlstad University, Region Värmland, and Sophiahemmet University.

Data Availability

The data that supports these study findings is available from the authors upon reasonable request.

Authors' Contributions

DAS, JN, UBJ, and AN were involved in designing the study, the questionnaire, and interview guide. Survey data was collected by DAS. Interviews were conducted by AN, while DAS acted as moderator and note taker. DAS and AN analyzed the data, which was critically reviewed by UBJ and JN and amended as needed. DAS drafted the first version of the manuscript, and along with AN, UBJ, and JN, critically reviewed and modified it. All 4 authors approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Brands and model of digital health technology and mobile health app features used by the participants.
[DOCX File, 20 KB - diabetes_v11i1e79704_app1.docx]

Multimedia Appendix 2
More examples of the data analysis process.
[DOCX File, 18 KB - diabetes_v11i1e79704_app2.docx]

References

1. Holt RIG, DeVries JH, Hess-Fischl A, et al. The management of type 1 diabetes in adults. a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2021 Nov;44(11):2589-2625. [doi: 10.2337/dci21-0043] [Medline: 34593612]
2. Wijk I, Amsberg S, Andreassen Gleissman S, Toft E, Anderbro T, Johansson UB. Living with type 1 diabetes as experienced by adults with prolonged elevated HbA1c: a qualitative study. Diabetes Ther 2023 Oct;14(10):1673-1684. [doi: 10.1007/s13300-023-01443-z] [Medline: 37470946]
3. Mobasseri M, Shirmohammadi M, Amiri T, Vahed N, Hosseini Fard H, Ghojazadeh M. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. Health Promot Perspect 2020;10(2):98-115. [doi: 10.34172/hpp.2020.18] [Medline: 32296622]
4. Ogle GD, James S, Dabelea D, et al. Global estimates of incidence of type 1 diabetes in children and adolescents: results from the International Diabetes Federation Atlas, 10th edition. Diabetes Res Clin Pract 2022 Jan;183:109083. [doi: 10.1016/j.diabres.2021.109083] [Medline: 34883188]
5. Harding JL, Wander PL, Zhang X, et al. The incidence of adult-onset type 1 diabetes: a systematic review from 32 countries and regions. Diabetes Care 2022 Apr 1;45(4):994-1006. [doi: 10.2337/dc21-1752] [Medline: 35349653]
6. American Association of Diabetes Educators. An effective model of diabetes care and education: revising the AADE7 Self-Care Behaviors. Diabetes Educ 2020 Apr;46(2):139-160. [doi: 10.1177/0145721719894903] [Medline: 31928334]
7. Liu Y, Liu C. Effect of the AADE7 Self-Care Behaviors Framework on diabetes education management in a shared care model. Int J Endocrinol 2024;2024:7278207. [doi: 10.1155/2024/7278207] [Medline: 38464361]
8. American Diabetes Association Professional Practice Committee. Diabetes technology: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S146-S166. [doi: 10.2337/dc25-S007] [Medline: 39651978]
9. Grunberger G, Sherr J, Allende M, et al. American Association of Clinical Endocrinology clinical practice guideline: the use of advanced technology in the management of persons with diabetes mellitus. Endocr Pract 2021 Jun;27(6):505-537. [doi: 10.1016/j.eprac.2021.04.008] [Medline: 34116789]
10. Pease A, Szwarcbard N, Earnest A, Andrikopoulos S, Wischer N, Zoungas S. Glycaemia and utilisation of technology across the lifespan of adults with type 1 diabetes: results of the Australian National Diabetes Audit (ANDA). Diabetes Res Clin Pract 2021 Jan;171:108609. [doi: 10.1016/j.diabres.2020.108609] [Medline: 33310120]
11. Penfornis A, Down S, Seignez A, Vives A, Bonnemaire M, Kulzer B. European survey on adult people with type 1 diabetes and their caregivers: insights into perceptions of technology. J Diabetes Sci Technol 2025 Mar;19(2):407-414. [doi: 10.1177/19322968231208690] [Medline: 37937589]
12. Kesavadev J, Saboo B, Krishna MB, Krishnan G. Evolution of insulin delivery devices: from syringes, pens, and pumps to DIY artificial pancreas. Diabetes Ther 2020 Jun;11(6):1251-1269. [doi: 10.1007/s13300-020-00831-z] [Medline: 32410184]
13. Didyuk O, Econom N, Guardia A, Livingston K, Klueh U. Continuous glucose monitoring devices: past, present, and future focus on the history and evolution of technological innovation. J Diabetes Sci Technol 2021 May;15(3):676-683. [doi: 10.1177/1932296819899394] [Medline: 31931614]
14. Speight J, Choudhary P, Wilmot EG, et al. Impact of glycaemic technologies on quality of life and related outcomes in adults with type 1 diabetes: a narrative review. Diabet Med 2023 Jan;40(1):e14944. [doi: 10.1111/dme.14944] [Medline: 36004676]
15. Rand L, Dunn M, Slade I, Upadhyaya S, Sheehan M. Understanding and using patient experiences as evidence in healthcare priority setting. Cost Eff Resour Alloc 2019;17(1):20. [doi: 10.1186/s12962-019-0188-1] [Medline: 31572067]
16. Natale P, Chen S, Chow CK, et al. Patient experiences of continuous glucose monitoring and sensor-augmented insulin pump therapy for diabetes: a systematic review of qualitative studies. J Diabetes 2023 Dec;15(12):1048-1069. [doi: 10.1111/1753-0407.13454] [Medline: 37551735]
17. Winterdijk P, Aanstoot HJ, Nefs G. The impact of real-time sensor technology on quality of life for adults with type 1 diabetes: a Dutch national survey. Diabetes Res Clin Pract 2023 Sep;203:110886. [doi: 10.1016/j.diabres.2023.110886] [Medline: 37604282]

XSL•FO
RenderX

18. Cleal B, Nexø MA, Schmidt S, Rytter K, Nørgaard K, Willaing I. Values and control in type 1 diabetes beyond glycemic outcomes: a qualitative interview study of everyday life with an insulin pump. Chronic Illn 2022 Sep;18(3):620-633. [doi: 10.1177/17423953211023962] [Medline: 34162271]
19. Wang LR, Malcolm J, Arnaout A, Humphrey-Murto S, LaDonna KA. Real-world patient experience of long-term hybrid closed-loop insulin pump use. Can J Diabetes 2021 Dec;45(8):750-756. [doi: 10.1016/j.jcjd.2021.02.006] [Medline: 33958309]
20. Garza KP, Howard KR, Feldman M, Weissberg-Benchell J. Adult's lived experience using the insulin-only bionic pancreas. J Diabetes Sci Technol 2025 Jan;19(1):11-17. [doi: 10.1177/19322968241274364] [Medline: 39212346]
21. Tanenbaum ML, Commissariat PV. Experience with burdens of diabetes device use that affect uptake and optimal use in people with type 1 diabetes. Endocr Connect 2023 Oct 1;12(10):e230193. [doi: 10.1530/EC-23-0193] [Medline: 37522857]
22. Anell A, Glenngård AH, Merkur S. Sweden health system review. Health Syst Transit 2012;14(5):1-159. [Medline: 22894859]
23. Socialdepartementet. Hälso- och sjukvårdslag (2017:30) [Report in Swedish]. Sveriges Riksdag. 2017. URL: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/halso--och-sjukvardslag-201730_sfs-2017-30 [accessed 2025-12-08]
24. Nationella diabetesregistret: årsrapport [Report in Swedish]. : Nationella Diabetesregistret; 2024 URL: https://registercentrum.blob.core.windows.net/ndr/r/NDR_-rsrapport-2024-nfsZyhug5.pdf [accessed 2025-12-08]
25. Stephen DA, Nordin A, Johansson UB, Nilsson J. Psychosocial self-efficacy and its association with selected potential factors among adults with type 1 diabetes: a cross-sectional survey study. Diabetes Ther 2024 Jun;15(6):1361-1373. [doi: 10.1007/s13300-024-01581-y] [Medline: 38642262]
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Stephen DA, Nordin A, Johansson UB, Nilsson J. eHealth literacy and its association with demographic factors, disease-specific factors, and well-being among adults with type 1 diabetes: cross-sectional survey study. JMIR Diabetes 2025 Mar 31;10:e66117. [doi: 10.2196/66117] [Medline: 40163628]
28. Fridlund B, Henricson M, Mårtensson J. Critical incident technique applied in nursing and healthcare sciences. SOJ Nurs Health Care 2017 Jan 23;3(1):1-5. [doi: 10.15226/2471-6529/3/1/00125]
29. Butterfield LD, Borgen WA, Amundson NE, Maglio AST. Fifty years of the critical incident technique: 1954-2004 and beyond. Qual Res 2005 Nov;5(4):475-497. [doi: 10.1177/1468794105056924]
30. Kemppainen JK. The critical incident technique and nursing care quality research. J Adv Nurs 2000 Nov;32(5):1264-1271. [doi: 10.1046/j.1365-2648.2000.01597.x] [Medline: 11115012]
31. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
32. Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: a discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
33. Elo S, Kääriäinen M, Kanste O, Pölkki T, Utriainen K, Kyngäs H. Qualitative content analysis: a focus on trustworthiness. SAGE Open 2014;4(1):2158244014522633. [doi: 10.1177/2158244014522633]
34. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2025-12-08]
35. Zahid M, Dowlatshahi S, Kansara AH, Sadhu AR. The evolution of diabetes technology - options toward personalized care. Endocr Pract 2023 Aug;29(8):653-662. [doi: 10.1016/j.eprac.2023.04.007] [Medline: 37100350]
36. Messer LH, Johnson R, Driscoll KA, Jones J. Best friend or spy: a qualitative meta-synthesis on the impact of continuous glucose monitoring on life with type 1 diabetes. Diabet Med 2018 Apr;35(4):409-418. [doi: 10.1111/dme.13568] [Medline: 29247556]
37. Burnside M, Haitana T, Crocket H, et al. Interviews with Indigenous Māori with type 1 diabetes using open-source automated insulin delivery in the CREATE randomised trial. J Diabetes Metab Disord 2023 Jun;22(1):861-871. [doi: 10.1007/s40200-023-01215-3] [Medline: 37250371]
38. Read M, Henshaw KN, Zaharieva DP, et al. Empowering us": a community-led survey of real-world perspectives of adults with type 1 diabetes using insulin pumps and continuous glucose monitoring to manage their glucose levels. Diabetes Res Clin Pract 2023 Aug;202:110830. [doi: 10.1016/j.diabres.2023.110830] [Medline: 37451626]
39. Persson M, Leksell J, Ernersson Å, Rosenqvist U, Hörnsten Å. 'Striving for freedom or remaining with what is well-known': a focus-group study of self-management among people with type 1 diabetes who have suboptimal glycaemic control despite continuous subcutaneous insulin infusion. BMJ Open 2022 Apr 1;12(4):e057836. [doi: 10.1136/bmjopen-2021-057836] [Medline: 35365536]
40. Sørgård B, Iversen MM, Mårtensson J. Continuous glucose monitoring in adults with type 1 diabetes: a balance between benefits and barriers: a critical incident study. J Clin Nurs 2019 Sep;28(17-18):3318-3329. [doi: 10.1111/jocn.14911] [Medline: 31090960]

XSL•FO
RenderX

41. Markov AM, Krutilova P, Cedeno AE, McGill JB, McKee AM. Interruption of continuous glucose monitoring: frequency and adverse consequences. J Diabetes Sci Technol 2024 Sep;18(5):1096-1101. [doi: 10.1177/19322968231156572] [Medline: 36824046]
42. Lorenzen JT, Madsen KP, Cleal B, et al. Associations between use of diabetes technology and diabetes distress: a Danish cross-sectional survey of adults with type 1 diabetes. BMJ Open 2024 Mar 25;14(3):e080053. [doi: 10.1136/bmjopen-2023-080053] [Medline: 38531585]
43. Barnard-Kelly KD, Martínez-Brocca MA, Glatzer T, Oliver N. Identifying the deficiencies of currently available CGM to improve uptake and benefit. Diabet Med 2024 Aug;41(8):e15338. [doi: 10.1111/dme.15338] [Medline: 38736324]
44. McKechnie V, Avari P, Eng PC, et al. The experiences of high-risk young adults with type 1 diabetes transitioning to real-time continuous glucose monitoring - a thematic analysis. PLoS ONE 2025;20(5):e0320702. [doi: 10.1371/journal.pone.0320702] [Medline: 40315203]
45. Rytter K, Hougaard A, Skouboe AG, Serifovski N, Ranjan AG, Nørgaard K. A new approach in insulin pump education improves glycemic outcomes: a randomized controlled trial. Acta Diabetol 2025 Jan;62(1):87-94. [doi: 10.1007/s00592-024-02340-y] [Medline: 39172172]
46. Patel PM, Thomas D, Liu Z, Aldrich-Renner S, Clemons M, Patel BV. Systematic review of disparities in continuous glucose monitoring and insulin pump utilization in the United States: key themes and evidentiary gaps. Diabetes Obes Metab 2024 Oct;26(10):4293-4301. [doi: 10.1111/dom.15774] [Medline: 39010293]
47. Doody O, Slevin E, Taggart L. Focus group interviews part 3: analysis. Br J Nurs 2013;22(5):266-269. [doi: 10.12968/bjon.2013.22.5.266] [Medline: 23545552]
48. Flanagan JC. The critical incident technique. Psychol Bull 1954 Jul;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
49. Addala A, Suttiratana SC, Wong JJ, et al. Cost considerations for adoption of diabetes technology are pervasive: a qualitative study of persons living with type 1 diabetes and their families. Diabet Med 2021 Oct;38(10):e14575. [doi: 10.1111/dme.14575] [Medline: 33794006]
50. Conway RB, Snell-Bergeon J, Honda-Kohmo K, et al. Disparities in diabetes technology uptake in youth and young adults with type 1 diabetes: a global perspective. J Endocr Soc 2024 Nov 26;9(1):bvae210. [doi: 10.1210/jendso/bvae210] [Medline: 39703363]

Abbreviations

AID: automated insulin delivery
CGM: continuous glucose monitoring
COREQ: consolidated criteria for reporting qualitative research
CSII: continuous subcutaneous insulin infusion
DHT: digital health technology
HbA_{1c}: glycated hemoglobin
mHealth: mobile health

Edited by A Stone, S Ashe; submitted 26.Jun.2025; peer-reviewed by G Ash, Y Shen; revised version received 09.Feb.2026; accepted 13.Feb.2026; published 26.Mar.2026.

Please cite as:
Stephen DA, Nilsson J, Johansson UB, Nordin A
Experiences of Adults With Type 1 Diabetes Using Digital Health Technology for Diabetes Self-Care: Qualitative Study
JMIR Diabetes 2026;11:e79704
URL: https://diabetes.jmir.org/2026/1/e79704
doi:10.2196/79704

© Divya Anna Stephen, Jan Nilsson, Unn-Britt Johansson, Anna Nordin. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 26.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Social Media Discussions of Anti-Diabetic Drugs and the Popularity of GLP-1 Therapies: Content Analysis

Chloe Stallion, BS, MPH; Tamkeen Khan, PhD; Christian Guerrero, MBA, MS; Stavros Tsipas, MA; Gregory Wozniak, PhD

American Medical Association, 330 N. Wabash Ave, Suite 39300, Chicago, IL, United States

Corresponding Author:
Chloe Stallion, BS, MPH
American Medical Association, 330 N. Wabash Ave, Suite 39300, Chicago, IL, United States

Abstract

Background: The clinical use of antidiabetic drugs (ADDs) has gained prominent public visibility due to the Food and Drug Administration (FDA) expansion of glucagon-like peptide-1 receptor agonists (GLP-1 RAs) in recent years; these drugs are now widely prescribed for weight loss. This shift is reflected in online public discourse.

Objective: This study aimed to analyze the current online public discourse surrounding ADDs, with an emphasis on perceptions of insulin versus noninsulin therapies and the increasing prominence of weight loss–associated medications.

Methods: We conducted a retrospective analysis of 23,580 English-language posts from the United States in 2023, using Boolean keyword searches to examine conversations about insulin-inclusive, with or without weight loss terms, and insulin-exclusive, with or without weight loss terms. We analyzed the volume and sentiment of online conversations regarding ADDs, frequency of drug mentions, proportion of posts by self-identified physicians, and thematic analysis of patient and clinician concerns.

Results: A total of 26,832 initial posts were collected in 2023. After post validation, 23,580 posts remained as the study sample size. Of the 23,580 posts, 4.6% (1087/23,580) mentioned insulin, and 95.4% (n=22,493) did not mention insulin. Semaglutide-containing drugs such as Ozempic and Mounjaro were the most referenced medications, particularly in weight-loss contexts. Weight loss conversations made up the majority of online posts. Conversations about insulin were marginal compared with conversations that did not mention insulin online. Only 7% to 10% of posts came from self-identified physicians. Key themes included drug accessibility, off-label use for weight management, concerns about supply shortages, insurance coverage, and growing calls for holistic care. Notably, public perspectives favored the dual efficacy of medications like Ozempic in managing diabetes and promoting weight loss.

Conclusions: The discourse on ADDs is dominated by weight loss–oriented therapies, with GLP-1 RAs driving much of the engagement. This surveillance paper illustrates the current state of public health interests around antidiabetic medications associated with weight loss and raises concerns regarding equitable access for patients with diabetes. These findings underscore the need for updated clinical guidance on necessary lifestyle behaviors for antidiabetic medication use and ongoing monitoring of public opinions regarding chronic disease management medications.

(JMIR Diabetes 2026;11:e87590) doi:10.2196/87590

KEYWORDS

diabetes; obesity; insulin; Twitter; social media; infodemiology; infoveillance; social listening; cost; rationing; weight loss; GLP

Introduction

Diabetes affects 38.1 million adults (14.7%) in the United States [1]. Diabetes is a significant risk factor for heart disease, stroke, and kidney problems [2].

Antidiabetic drugs (ADDs) are vital for managing type 2 diabetes by maintaining blood glucose levels [3]. Traditionally, insulin-based therapies were the mainstay; however, recent years have seen a surge in noninsulin options, such as glucagon-like peptide-1 receptor agonists (GLP-1 RAs) [1,3,4]. The weight loss side effect associated with taking these drugs has made certain medications a consideration for patients who struggle to lose weight but do not have type 2 diabetes [5]. In particular, 2 GLP-1 RA molecules, semaglutide and tirzepatide, are indicated for type 2 diabetes but have also been approved for weight loss, under different brand names [6,7]. Semaglutide was Food and Drug Administration (FDA)–approved to treat type 2 diabetes in 2017 [8]. These medical advancements in weight loss have become major topics within the diabetes conversation online. Physicians also prescribe semaglutide off-label to patients to help promote weight loss, which has increased the demand for diabetes medications [9-11]. With FDA approval of tirzepatide to treat type 2 diabetes in the spring

XSL•FO
RenderX

of 2022, there was a 400% increase in ADD prescriptions by nonendocrinologists, and the drug was subsequently approved for weight loss [7,12].

Earlier studies have explored the public perception of ADDs, showing an overall positive opinion driven by diabetes programs and patients [13]. With more recent approvals for specific GLP-1 RAs to be used for weight management, there is an increased demand for these medications for reasons other than diabetes management [12,14]. Total spending on GLP-1 RAs increased by more than 500% from 2018 to 2023 [15], and state-level spending and fills were higher in states with higher obesity prevalence [16]. This study aimed to analyze the notable online discourse surrounding ADDs, emphasizing the growing preference for weight loss–associated therapies over traditional insulin-based treatments as seen in online conversations. Antidiabetic medications have become a popular topic of interest, and monitoring online conversation shows that these medications are largely discussed in association with weight loss in nondiabetic patients. As a surveillance study, our aim was to monitor and report on the current online discourse surrounding antidiabetic medications.

Methods

Overview

This methodology provides insight into the current use of antidiabetic medications available in the marketplace. Although this study did not evaluate causal mechanisms, observations of the most frequently mentioned drugs may be associated with factors such as the availability of pharmacological options, direct-to-consumer marketing, perceived efficacy, and affordability. The methodological approach applied in this surveillance study offers a foundation for future research to systematically assess these potential influences. We created a query to collect all publicly available posts posted between January 1, 2023, and December 31, 2023, on X (formerly Twitter). The methodology involved developing a structured query using a targeted set of keywords and phrases related to the study of ADD use. These terms were entered in various combinations, accounting for variations in spelling, brand and generic drug names, and contextual keywords to improve precision. There were 2 distinct queries: one designed to capture conversation around insulin, and another excluding insulin. This approach enabled a nuanced analysis of perspectives, themes, and engagement patterns across different audience segments.

Query Structure

The query structure shown in Multimedia Appendix 1 has 2 main queries: query 1 and query 2, which mirror the previous study [13]. Query 1 explores ADDs conversations including insulin, and query 2 explores ADDs conversations excluding insulin. This updated study expands each of those queries by adding a subquery that differentiates conversations including weight loss and conversations excluding weight loss. Query 1a in query 1 of the study explores ADDs and insulin but excludes weight loss–associated ADDs; query 1b in query 1 of the study explores only weight loss–associated ADDs and insulin conversations. In contrast, query 2a in query 2 of the study explores ADDs, excluding insulin, but includes weight loss–associated ADDs; query 2b in query 2 explores only weight loss–associated ADDs and excludes insulin. The general keywords related to diabetes and insulin were kept consistent with the previous study, using Boolean language to include or exclude insulin. The ADD terms from the previous study were separated into those not associated with weight loss and those associated with weight loss and were used for query 1b and query 2b, respectively.

Keyword Selection

Keywords were selected using the query terms, which included search terms associated with diabetes, insulin, and ADDs (Multimedia Appendix 2). ADDs are identified by their generic and brand names. Certain drugs are sold under one FDA-approved name, and others are sold under multiple FDA-approved brand names; therefore, the search terms included generic names, brand names, and common misspellings. (Multimedia Appendix 3). Expanding upon the previous study, this study aimed to explore differences in online conversations between drugs not associated and those positively associated with weight loss. Additional terms added in this study include *BMI and* A_{1c} to further separate the data into those concerning and not concerning weight loss; however, because health care professionals largely use these terms, they excessively limited the scope of the query when used as exclusionary terms. Instead, they were used as supplementary terms for diabetes.

Query Filtering

Topic modeling was conducted on each query to identify noise that entered the initial results and did not pertain to the topic outlined in the study. Keywords associated with the noise were excluded from the query results, which yielded a body of posts within each query that focused on antidiabetic medication use without noise.

The posts were filtered to remove results from protected (private) X accounts, in which post content data were not visible. Additional filters removed posts originating outside the United States and those in languages other than English. Original posts were the primary focus for understanding the online conversation; therefore, reposts and quote posts were excluded from the query. To best capture the breadth of the conversation, only original posts were included. Finally, 8 noise categories—advertisements, autoposts, contests and games, financial news, scams, job offers, not safe for work, and vulgar content—were excluded from the query to limit irrelevant conversation.

Descriptive Data Analysis

ADD Counts

Drug counts were created by calculating the number of mentions of each drug across every query. In the online space, both the chemical names and the brand names are used to identify ADDs. The chemical compounds and their brand-name counterparts were combined for each drug count, except for semaglutides, since compounded semaglutide is distinct from brand-name semaglutides. Different drug terms were used within each query at varying rates, highlighting which medications were more

XSL•FO
RenderX

widely used and sought after than others. Brand names and generic names for these medications were used interchangeably online; therefore, generic drugs that were sold under a specific FDA-approved brand name were grouped as one term (eg, Victoza and liraglutide). Conversely, generic drug names that are sold under multiple FDA-approved brand names are counted separately (eg, Mounjaro, Zepbound, and tirzepatide).

Physician Counts

Posts made by clinicians were identified by "md" and "medical doctor" keywords in the author's name and username (Multimedia Appendix 4). The term "doctor" in isolation was not used as a query keyword because it encompasses far too large a range of nonmedical professions (eg, PhD). Variations of the abbreviated forms of "medical doctor" were used to account for common spellings. Misspellings of "medical doctor" or "med doc" were not included. These posts were manually filtered to remove false-positive individuals and medical organizations as this study focuses on clinicians (ie, physicians).

Qualitative Content Analysis

A proprietary natural language processing algorithm created by Ipsos using Google BERT's AI language model was used to categorize opinions and perspectives [17]. Physician-related posts were identified through keywords indicating medical credentials. These exclusion criteria allowed for a more precise manual evaluation of authentic conversations by ensuring that only organic, user-generated discussions were analyzed. Manual analysis was performed to group themes for each query and subquery within the study.

Ethical Considerations

Because no individuals were put at risk during this data collection process, this study did not meet the requirements to be considered human subjects research. The data for this project were collected through a web scraping of public X data powered by Ipsos and did not require any interaction with human participants. We observed and aggregated publicly available data. The American Medical Association partnered with the University of Illinois-Chicago for institutional review board processes. Their determination is provided below:

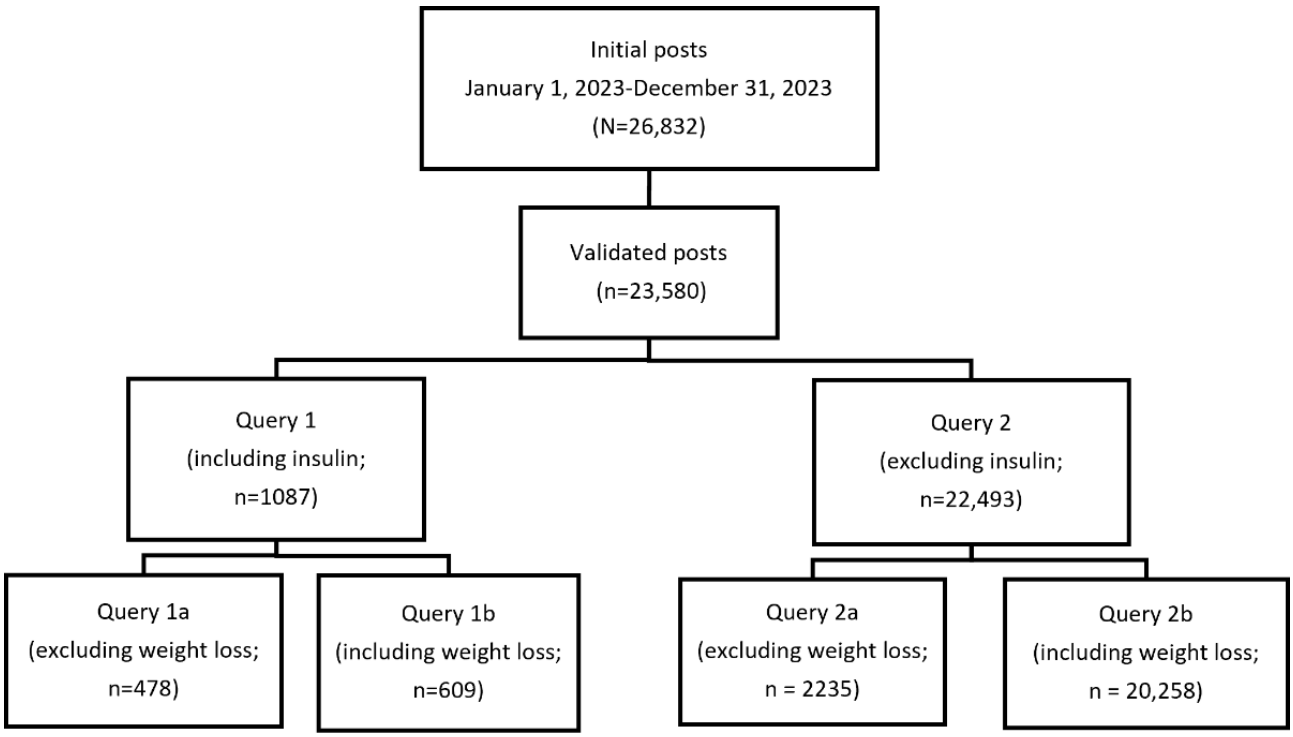
> *Because you are not engaging with human subjects and all of the data being analyzed is publicly available information, this research would not qualify as human subjects research.*

Results

Volume of Posts

The number of posts for the 4 queries varied over the observation period, as shown in Figure 1, with more insulin-related diabetes medication conversations discussing weight loss than conversations that did not discuss weight loss. Query 2 (excluding insulin conversation) has a much larger number of posts than query 1, encompassing 95.4% (22,493/23,580) of the validated posts, as seen in Figure 1. Within query 2, the overwhelming majority (20,258/22,493, 90%) of the posts discussed weight loss. Noninsulin ADDs associated with weight loss made up most of the diabetic conversations online. This suggests that most of the conversations online surrounding diabetes do not include insulin-related treatments and instead focus on ADDs that are associated with weight loss. These results illustrate that weight loss–associated with ADDs currently dominated online diabetes conversations.

Figure 1. Flowchart of post validation and categorization.

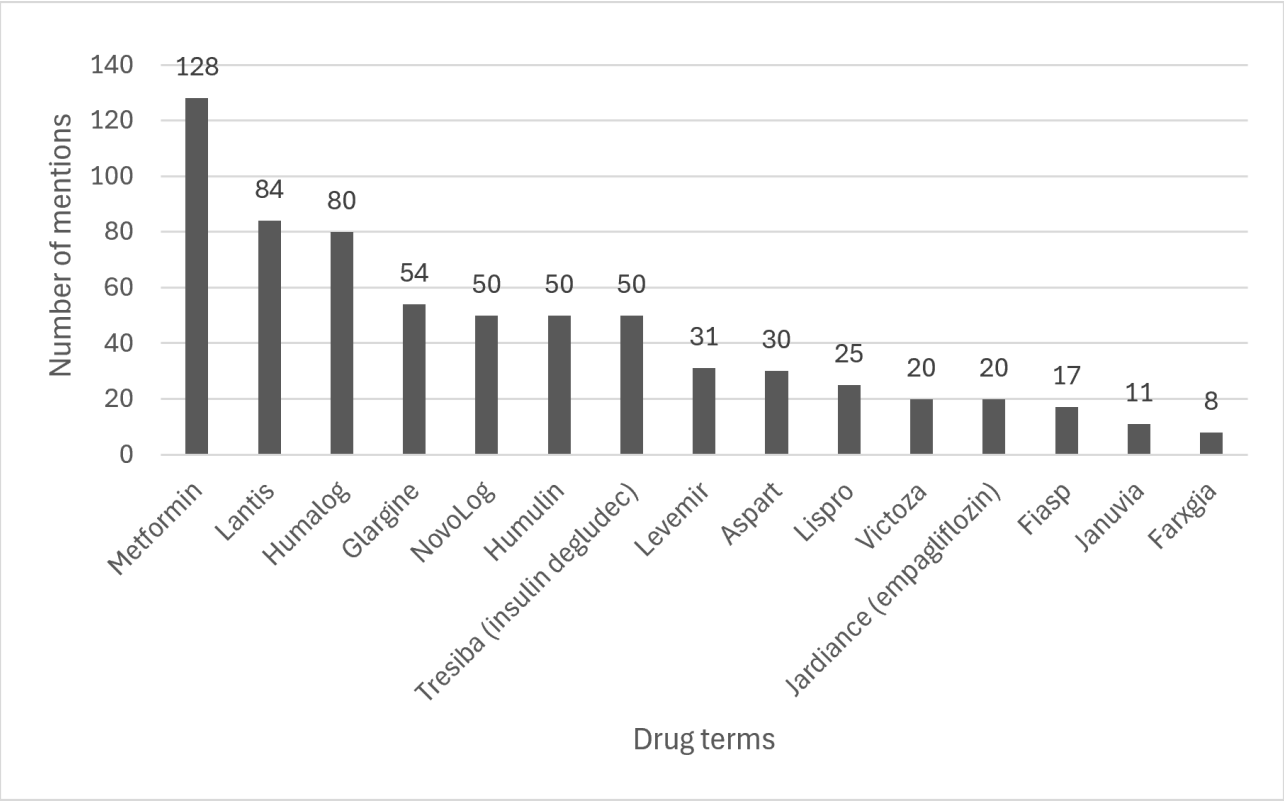


XSL•FO
RenderX

Drug References

Figure 2 shows query 1a (including insulin and excluding weight loss) contained 39 specific ADDs. Of those ADDs, metformin (n=128) had the most references online, followed by Lantus (n=84), a long-acting insulin product, which had the next most references. Humalog (n=80) and Novolog (n=50), both rapid-acting insulin medications, followed in online popularity. All the drugs listed above can be used to treat type 1 and type 2 diabetes, except metformin, which is FDA-approved to treat only type 2 diabetes. Metformin was mentioned in 27% (128/478) of the posts in query 1a.

Figure 2. Top 15 drug terms mentioned in query 1a.



As shown in Figure 3, among the query 1b (including insulin and weight loss) online conversations, the most popular ADD mentioned was Ozempic (n=119). Semaglutides (n=60), the generic name for a GLP-1 receptor agonist, was the next most frequently mentioned drug online. Even though semaglutide and Ozempic share the same active ingredient, there are compounded versions of these drugs (compounded semaglutides) that are not FDA-approved and have been shown to cause medical complications for patients who use them, posing a concern for physicians [5,6,18]. Ozempic was used twice as often in this online discussion as the generic term semaglutide. The remaining drugs were mentioned less than half the number of times as Ozempic and semaglutides. Metformin (n=24), Mounjaro (n=19), and Trulicity (n=14) were mentioned as the next highest in this query. Ozempic was mentioned in 20% (119/609) of the posts in query 1b. When weight loss is included in the online ADD conversation, Ozempic and related semaglutides were the ADDs most frequently discussed online.

Figure 4 shows query 2a (excluding insulin and weight loss), metformin (n=1145) was still the most referenced antidiabetic medication. While not as heavily associated with weight loss as other antidiabetic medications, metformin has been shown to lessen weight gain in patients taking certain antipsychotic medications [19]. Jardiance (empagliflozin; n=563) was the next highly referenced antidiabetic medication used to treat type 2 diabetes. Farxiga (dapagliflozin; n=115) was the third highest referenced drug, followed by Januvia (n=88). Of these medications, Farxiga was the only medication approved to treat more than type 2 diabetes—it could also be used to prevent heart failure and chronic kidney disease [20]. Metformin was mentioned approximately 512.3 times per 1000 posts, meaning that almost half of the 2235 posts in query 2a mentioned metformin.

XSL•FO
RenderX

Figure 3. Top 14 drug terms mentioned in query 1b.

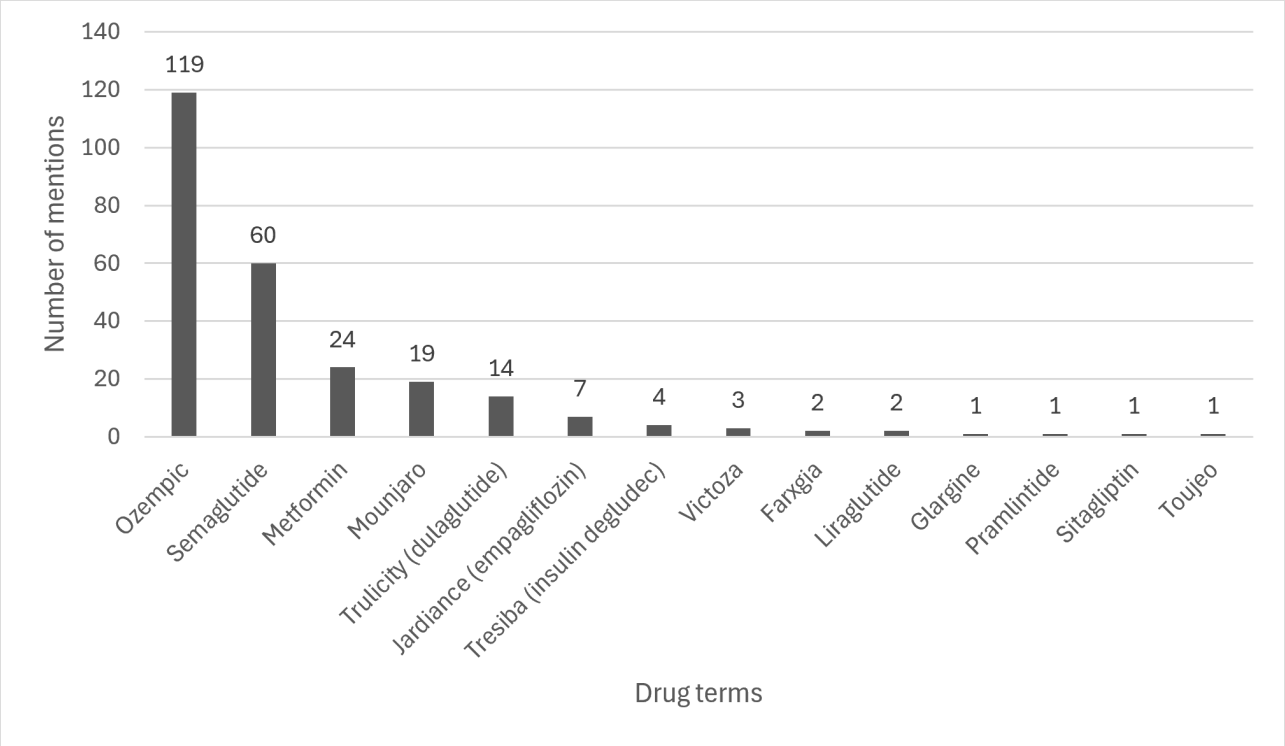
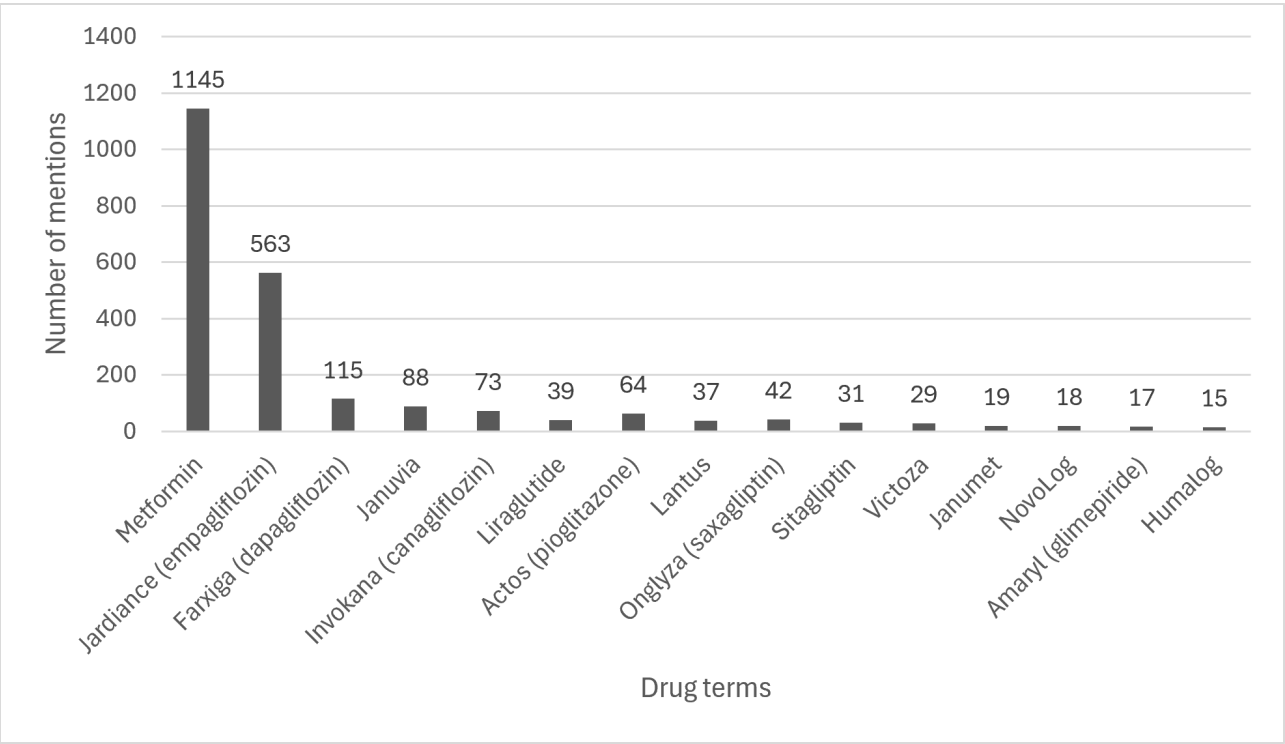


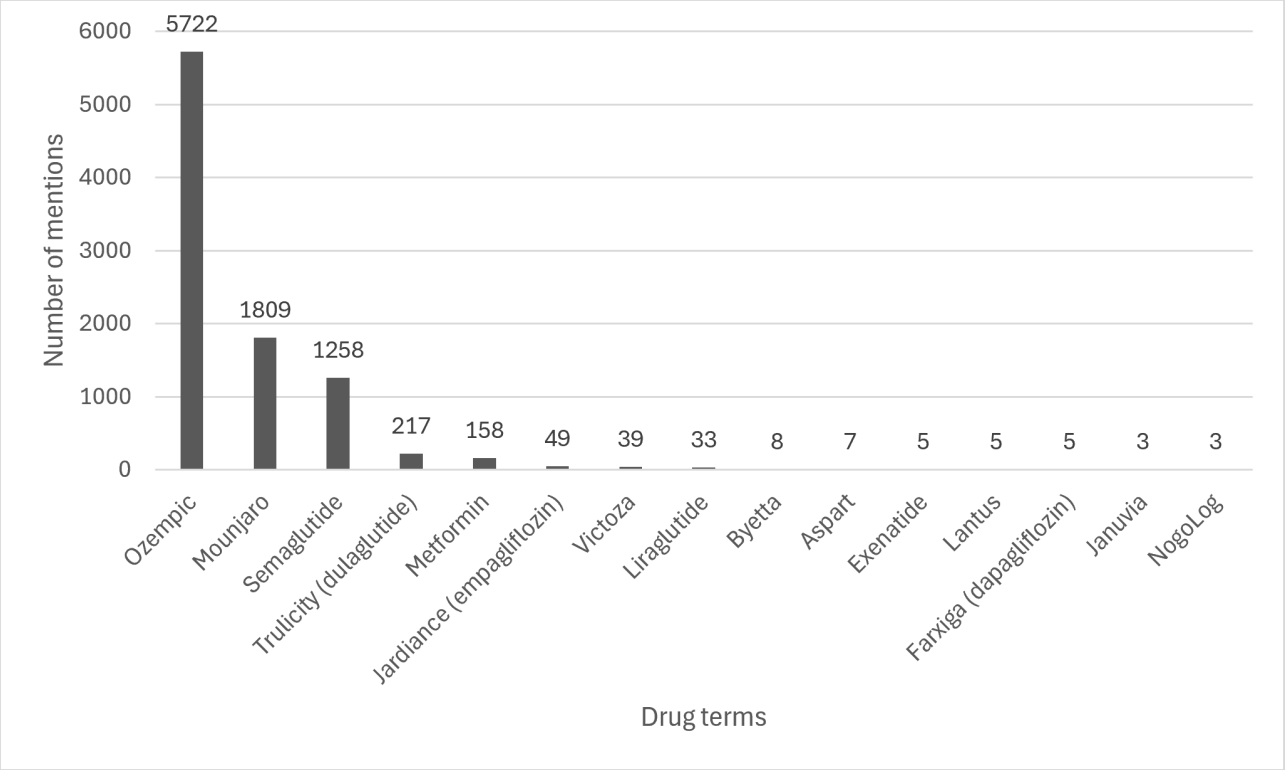
Figure 4. Top 15 drug terms mentioned in query 2a.



Within query 2b, Ozempic (n=5722) was the drug most mentioned, appearing in 28% (5722/20,258) of the posts as seen in Figure 5. The next most discussed ADD was Mounjaro (n=1809), appearing in 9% (n=1809) and semaglutides (n=1258) appearing in 6% (n=1258) of the posts in query 2b. Ozempic was mentioned approximately 282.5 times per 1000 posts, meaning that 30% of the posts in query 2b were discussing Ozempic. Brand names such as Ozempic, versus its generic name, semaglutide, are used interchangeably online; however, the brand name was used 4.5 times more online than the generic name. The greater recognition of Ozempic in online discussions suggests that patients and other individuals may be more familiar with Ozempic than with semaglutide.

XSL•FO
RenderX

Figure 5. Top 15 drug terms mentioned in query 2b.



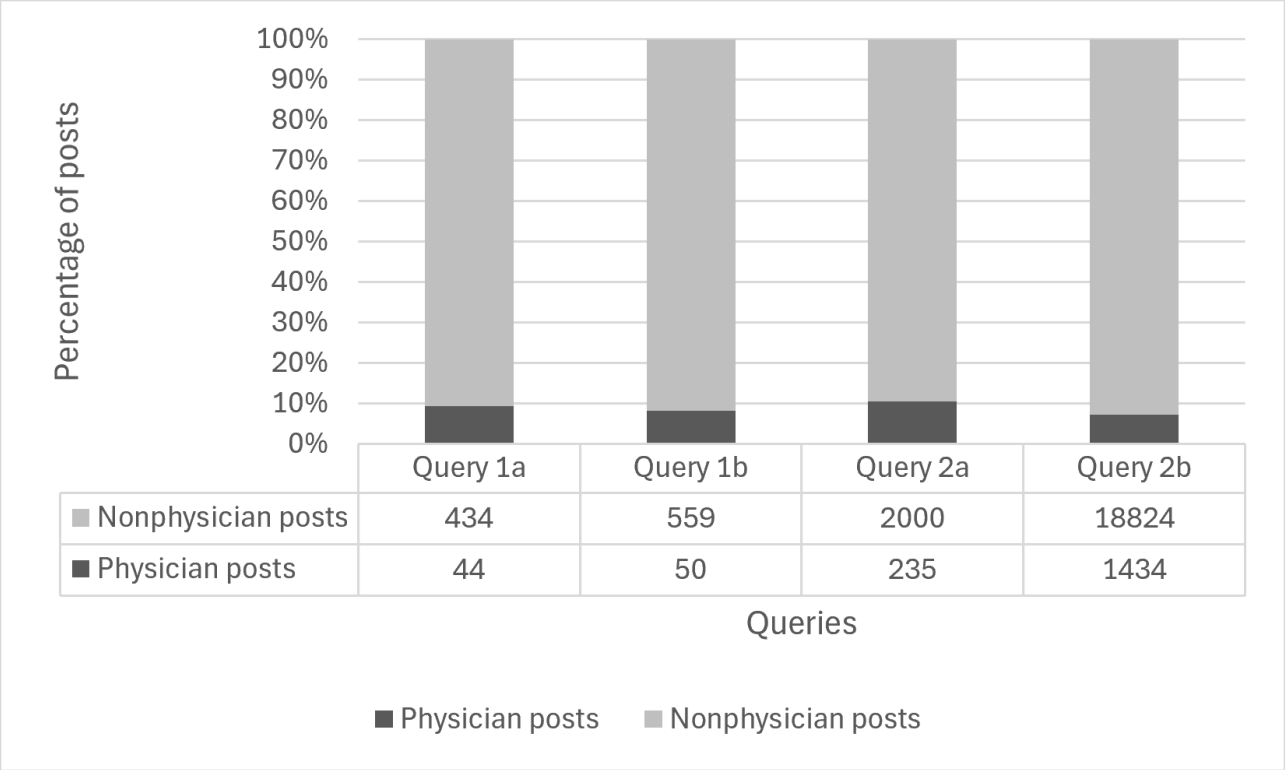
When insulin was included in the online ADD conversation (query 1), metformin was the most popular drug referenced in query 1. In contrast, when weight loss was included in the conversation, regardless of insulin, Ozempic became the most frequently mentioned drug, surpassing metformin in the number of references. This finding underscores the noteworthy influence Ozempic has on the current ADD market.

Physician Count

Within query 1a (including insulin and excluding weight loss), 9.2% (44/478) of posts came from self-identified physicians. Within query 1b (including insulin and weight loss), 8.2% (50/609) of posts came from self-identified physicians. Query 2a (excluding insulin and weight loss) contained 10.5% (235/2235) of posts from self-identified physicians. Query 2b (excluding insulin and including weight loss) had 7.1% (1434/20,258) of the online conversation. Within each of the 4 query subsections, clinicians made up at most 10.5% of the conversation, meaning that most of the online conversation was not by medical professionals, as seen in Figure 6.

XSL•FO
RenderX

Figure 6. Percentage of posts by self-identified physicians.



Thematic Analysis

Query 1a focused on ADD-related posts that included insulin and excluded weight loss. Several key points emerged from this subset of the conversation. Using Synthesio's Gen AI Trend Clusters capability, several prominent themes were identified, including insulin therapy, which is the gold standard for treating type 1 diabetes and more severe cases of type 2 diabetes. This involves regular insulin injections to help maintain proper body function. There was growing interest in online conversations about more innovative ways to administer insulin, such as patches or inhalation, which are less strenuous and taxing on the body. There was also interest in innovative insulin options that require only weekly injections rather than daily ones.

Individuals expressed concerns about access to affordable insulin treatments, including over-the-counter options that do not require a physician's prescription. Consumers viewed over-the-counter options as more affordable and accessible than prescription alternatives. Additionally, there was a growing demand for preventative antidiabetic measures, such as improved access to healthy food, exercise, and better living habits, along with stronger patient advocacy to prevent hyperglycemia and type 2 diabetes. Many were concerned that the health care industry prioritized treating diabetes rather than focusing on preventing its onset.

In contrast, query 1b also included insulin but added weight loss to the discussion. Both query 1a and 1b highlighted significant conversations around innovative and less invasive antidiabetic treatments, particularly noninjectables such as metformin. Because weight loss was a focus, these discussions centered less on the cost of and access to medications and more on the desire for access to healthier lifestyles and physical well-being. Additionally, there was a strong interest in incorporating community support opportunities into antidiabetic treatment plans. Among the innovative treatment options, both patients and clinicians advocated for more digital health tools, enabling physicians to remotely monitor patients' lifestyle behaviors and provide virtual care.

Query 2a excluded both insulin and weight loss posts from the online diabetes conversation, resulting in the emergence of vastly different themes. Notably, holistic medicine and personalized care were prominent topics, as many patients preferred less invasive, individualized treatments tailored to their specific diabetic symptoms. These treatments included vitamin and mineral supplementation, plant-based diets, integrated care (combining medicinal and holistic approaches), education, and community support groups. This trend contrasts sharply with conversations that include insulin and underscores the desire among many diabetic and prediabetic individuals for lifestyle-focused treatments aimed at preventing the symptoms associated with the disease. Finally, this query highlighted the impact of diabetes on mental health and the growing demand for treatments that incorporated psychotherapy.

Query 2b excluded insulin and included weight loss posts, which had the largest post count of all the queries. Within this query, patients highlighted the effectiveness of drugs such as Ozempic in treating their diabetic symptoms as well as the weight loss benefits this drug is associated with. Similarly, patients and health care providers agreed that strict diet and exercise routines increase the efficacy of these medications. Additionally, the demand for Ozempic and similar medications is rapidly increasing as they can now be used to primarily treat weight loss, not only diabetes. The demand for these drugs has brought vocalized disdain from diabetic patients about the strain this

XSL•FO
RenderX

puts on companies to keep these medications in stock for diabetic patients. There is also the fear that the increase in popularity of weight loss–associated drugs will increase the price of the medication and decrease insurance coverage. Diabetic patients worry that prior authorization will pose issues in receiving the drugs in the future, and physicians and health care professionals lament unaffordable prices for drugs such as Ozempic and Mounjaro in the United States in comparison to other countries.

Physician Themes

Clinician input was manually analyzed by focusing on posts from users who self-identified as physicians through terms such as MD, doctor, physician, or clinician in their name or bio. This small subset of posts highlighted several important conversations and concerns physicians had regarding antidiabetic medications.

In query 1a, posts from self-identified physicians made up 9.2% (44/478) of the results. This group praised the use of insulin treatments requiring only weekly injections instead of daily ones and supported the increased use and availability of noninsulin medications for treating type 2 diabetes. Additionally, clinicians emphasized the importance of healthy lifestyles, including proper nutrition and exercise, as lifestyle choices significantly affected obesity and type 2 diabetes symptoms. Query 1b, making up 8.2% (50/559) of the posts, revealed that physicians associated weight loss with the mitigation of type 2 diabetes symptoms, and that Ozempic could aid in this process. Additionally, clinicians believe health care resources should prioritize the prevention of type 2 diabetes through lifestyle behaviors rather than its management after onset.

In contrast, within query 2a, which comprised 10.5% (235/2000) of the posts, physicians highlighted both the positive and negative aspects of metformin. While metformin offers benefits beyond insulin regulation for type 2 diabetes patients, such as stabilizing white blood cell counts, reducing inflammation, preventing cancer, prolonging lifespan, and supporting proper immune responses, physicians also cautioned that discontinuing metformin in geriatric patients with diabetes may accelerate cognitive decline.

Query 2b mirrors query 1b, with physicians making up 7.1% (1434/18,824) of the conversation. As in query 1b, physicians consistently emphasized the importance of combining dietary and active lifestyle behaviors with innovative medications. Uniquely, query 2b also highlighted how Ozempic and similar GLP-1 RA drugs have revolutionized the antidiabetic treatment experience. When paired with proper nutrition and exercise, these medications help patients lose weight and effectively manage diabetic symptoms.

Discussion

Principal Findings

Our findings reveal that within the diabetic medication landscape, there is a notable focus on insulin alternatives such as Ozempic, reflecting their current medical popularity. This surveillance study revealed that conversations excluding insulin made up 95% of the diabetes conversation online, meaning insulin accounted for a small proportion of current public discourse on diabetes. Topics within this insulin-excluded conversation predominantly revolve around weight loss–associated drugs, which offer additional benefits beyond insulin regulation, enhancing their appeal. This surveillance study also revealed a current focus in the health care market: laypersons—both patients and nonpatients—dominate the conversation on ADDs, whereas physicians contribute only a small portion. FDA-approved GLP-1 RA medications such as Ozempic and Mounjaro are prescribed by physicians to treat weight loss [9]. This current pharmaceutical landscape is reflected online, and this study monitors and highlights this current public interest in ADDs for treating weight loss in patients without diabetes [21]. Wellness advocates, diabetes advocacy groups, and health-related accounts also significantly contribute, often emphasizing the potential risks and considerations for those using GLP-1 RAs such as Ozempic. Unlike insulin-based treatments, these drugs offer additional benefits beyond managing insulin levels, contributing to their rising popularity. Furthermore, when weight loss is included in the conversation online, Ozempic and semaglutides are mentioned far more frequently than any other ADD brand name. When weight loss is part of the discourse, mentions of Ozempic and semaglutides surpass those of other ADDs. This suggests a significant public interest in ADDs for nondiabetic use. The findings underscore the evolving role of weight loss associations in ADD discussions, reflecting the growing use of these medications for weight management in nondiabetic patients, both as Wegovy and through off-label prescriptions. This strongly indicates that online discussions about ADDs, excluding insulin, often focus on drugs associated with weight loss. This study elucidates the current popularity of ADDs in online conversation as a microcosm for our medical and health care landscape.

Although X (formerly Twitter) is a single platform and is not comprehensive of the entire US population, its robust usership within the United States makes the platform a powerful tool for observing public discourse. It is imperative that further studies be conducted across various prominent social media platforms to expand the knowledge on this emerging public health issue. Because this study did not use demographic data, it did not qualify as human subjects research; therefore, demographic data are not available to represent characteristics of the average user. As an observational study, trend analyses were not used; rather, the study focused on actual counts to measure the prevalence of the online discussion. This study used X data available in 2023, and future studies will continue to survey the progression of the ADD landscape.

Limitations

It is important to note that although this study provides valuable insights that add to the previous literature, limitations still exist. First, the queries used did not isolate the term “semaglutide” from “compounded semaglutide.” A manual analysis of the posts found that semaglutide and compound semaglutides were used interchangeably by patients and laypersons online. Therefore, the counts of semaglutides mentions in 2023 include both “compounded semaglutide” and “semaglutide” references.

XSL•FO
RenderX

Additionally, some clinicians may have been excluded from the study because they used “doctor” in their social media bio or username, which was not included as a keyword in this study to avoid including individuals with PhD degrees in the conversation results as noise. Therefore, medical doctors who self-identified on social media only as “doctor” without “md” or “medical” also in their bio would not have been captured in the physician sample set.

Furthermore, the data were constrained to English-language posts; therefore, this study could only survey and report on patients and individuals who speak and write English, excluding non–English-speaking medical populations.

Finally, these surveillance results show the emerging landscape of these medications and their rise in popularity. The underlying mechanisms may include, but are not limited to, increased pharmacological options, direct-to-consumer marketing, efficacy, and affordability. This surveillance study provides a foundation for future research to better understand these mechanisms.

Conclusions

This study highlighted the current popularity of ADDs, specifically GLP-1 RAs, in the online diabetic medication conversation. GLP-1 RA medications dominate conversation (~23,000 posts), and these medications were the most referenced for treating diabetes. Online laypersons expressed apprehension about accessing these medications for purposes other than diabetes treatment, causing concern among diabetic patients, who voiced these concerns online. Many diabetics are worried that the exponential increase in demand for ADDs for nondiabetic users for cosmetic purposes would increase the burden on supply and cause shortages or price hikes for those who need these medications to maintain body homeostasis [22]. Ozempic was originally FDA-approved to treat type 2 diabetes but has been prescribed off-label and was recently approved under the name Wegovy to promote weight loss in patients with obesity [8,18]. These federal approvals mark a key moment in the health care industry, reflecting a popularity in ADDs use. Additionally, laypersons used the term “semaglutide” and “compounded semaglutide” interchangeably online, often using the term “semaglutide” to refer to the non–FDA-approved compounded forms of these medications used to treat weight loss, which could cause some confusion about the actual drug being referred to, the former being FDA approved and the latter not FDA approved. Although the terms were used interchangeably, Ozempic was used overwhelmingly more often in online diabetic conversations than the generic term semaglutide. This evidence offers an insight into the fact that individuals and patients were more aware of Ozempic and its uses than of semaglutides. Physicians primarily used the term “compounded semaglutide” to voice their concern over patients going to medical spas to receive non–FDA-approved medication to try and mimic the weight loss effects and putting their health at risk to achieve weight loss. The FDA has published work supporting this online concern by noting that there have been anecdotal physician reports of patients experiencing gastroparesis as a result of using compounded semaglutides [6]. Physicians noted that antidiabetic medications are profoundly serious drugs that require changes in lifestyle and behavior to have the most positive effects during and after treatment.

Acknowledgments

The content is solely the responsibility of the authors and does not represent the official views of the American Medical Association. The study team wishes to acknowledge and thank Annalyn Skipper for her review of the manuscript before submission.

Funding

The authors declare no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Query structure.
[PNG File, 53 KB - diabetes_v11i1e87590_app1.png]

Multimedia Appendix 2
Query terms.
[DOCX File, 15 KB - diabetes_v11i1e87590_app2.docx]

Multimedia Appendix 3
Noninsulin diabetic medications and misspellings.
[DOCX File, 14 KB - diabetes_v11i1e87590_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Self-identified physician post isolation.
[DOCX File, 13 KB - diabetes_v11i1e87590_app4.docx]

References

1. National diabetes statistics report. : Centers for Disease Control and Prevention; 2026 URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2026-07-26]
2. Diabetes - symptoms and causes. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444 [accessed 2026-07-26]
3. Diabetes basics. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/about/index.html [accessed 2026-07-26]
4. Oral & injectable medications for type 2 diabetes. American Diabetes Association. URL: https://diabetes.org/health-wellness/medication/oral-other-injectable-diabetes-medications [accessed 2026-07-26]
5. Drug trials snapshots: MOUNJARO. U.S. Food and Drug Administration. 2023. URL: https://www.fda.gov/drugs/drug-approvals-and-databases/drug-trials-snapshots-mounjaro [accessed 2026-07-26]
6. FDA's concerns with unapproved GLP-1 drugs used for weight loss. U.S. Food and Drug Administration. URL: https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/medications-containing-semaglutide-marketed-type-2-diabetes-or-weight-loss [accessed 2026-07-26]
7. FDA approves new medication for chronic weight management. U.S. Food and Drug Administration. 2023. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-new-medication-chronic-weight-management [accessed 2026-07-26]
8. FDA approves first treatment to reduce risk of serious heart problems specifically in adults with obesity or overweight. U.S. Food and Drug Administration. 2024. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-reduce-risk-serious-heart-problems-specifically-adults-obesity-or [accessed 2026-07-26]
9. Glass R, Foster E. New demand in an old market. IQVIA. 2023. URL: https://www.iqvia.com/locations/united-states/blogs/2023/08/new-demand-in-an-old-market [accessed 2026-07-26]
10. Han SH, Safeek R, Ockerman K, et al. Public interest in the off-label use of glucagon-like peptide 1 agonists (Ozempic) for cosmetic weight loss: a Google Trends analysis. Aesthet Surg J 2023 Dec 14;44(1):60-67. [doi: 10.1093/asj/sjad211] [Medline: 37402640]
11. Rodriguez PJ, Goodwin Cartwright BM, Gratzl S, et al. Semaglutide vs tirzepatide for weight loss in adults with overweight or obesity. JAMA Intern Med 2024 Sep 1;184(9):1056-1064. [doi: 10.1001/jamainternmed.2024.2525] [Medline: 38976257]
12. Scannell C, Romley J, Myerson R, Goldman D, Qato DM. Prescription fills for semaglutide products by payment method. JAMA Health Forum 2024 Aug 2;5(8):e242026. [doi: 10.1001/jamahealthforum.2024.2026] [Medline: 39093590]
13. Golder S, Bach M, O'Connor K, Gross R, Hennessy S, Gonzalez Hernandez G. Public perspectives on anti-diabetic drugs: exploratory analysis of Twitter posts. JMIR Diabetes 2021 Jan 26;6(1):e24681. [doi: 10.2196/24681] [Medline: 33496671]
14. Shmerling RH. What happens when a drug goes viral? Harvard Health Publishing. 2023. URL: https://www.health.harvard.edu/blog/what-happens-when-a-drug-goes-viral-202302212892 [accessed 2026-07-26]
15. Tsipas S, Khan T, Loustalot F, Myftari K, Wozniak G. Spending on glucagon-like peptide-1 receptor agonists among US adults. JAMA Netw Open 2025;8(4):e252964. [doi: 10.1001/jamanetworkopen.2025.2964]
16. Khan T, Tsipas S, Loustalot F, et al. State-level variation in glucagon-like peptide-1 receptor agonists fills and spending by obesity prevalence, 2023. J Am Heart Assoc 2025 Aug 5;14(15):e043137. [doi: 10.1161/JAHA.125.043137] [Medline: 40673513]
17. Automated sentiment analysis (ASA). Ipsos. URL: https://resources.ipsos.com/rs/297-CXJ-795/images/Synthesio%27s%20Automated%20Sentiment%20Analysis%20Model.pdf [accessed 2026-07-26]
18. Callaghan T, Motta M, Stein M, Goidel K. The prevalence and consequences of support for off-label Ozempic prescriptions. Health Econ Policy Law 2025 Jan 7:1-14. [doi: 10.1017/S1744133124000306] [Medline: 39763244]
19. Shmerling RH. Is metformin a wonder drug? Harvard Health Publishing. 2024. URL: https://www.health.harvard.edu/blog/is-metformin-a-wonder-drug-202109292605 [accessed 2026-07-26]
20. Dapagliflozin tablets. Cleveland Clinic. URL: https://my.clevelandclinic.org/health/drugs/19382-dapagliflozin-tablets [accessed 2026-07-26]
21. Rodriguez P, Pantalone KM, Griebeler ML, Burguera B. Should I consider metformin therapy for weight loss in patients with obesity but without diabetes? Cleve Clin J Med 2023 Sep 1;90(9):545-548. [doi: 10.3949/ccjm.90a.22096] [Medline: 37657834]
22. Middleton J. Semaglutide shortage highlights inequities in diabetic medication access. American Family Physician. 2023. URL: https://www.aafp.org/pubs/afp/afp-community-blog/entry/semaglutide-shortage-highlights-inequities-in-diabetic-medication-access.html [accessed 2026-07-26]

Abbreviations

ADD: antidiabetic drug

XSL•FO
RenderX

FDA: Food and Drug Administration
GLP-1 RA: glucagon-like peptide 1 receptor agonist

Edited by R Zhao; submitted 11.Nov.2025; peer-reviewed by H Wu; revised version received 29.May.2026; accepted 31.May.2026; published 26.Aug.2026.

Please cite as:
Stallion C, Khan T, Guerrero C, Tsipas S, Wozniak G
Social Media Discussions of Anti-Diabetic Drugs and the Popularity of GLP-1 Therapies: Content Analysis
JMIR Diabetes 2026;11:e87590
URL: https://diabetes.jmir.org/2026/1/e87590
doi:10.2196/87590

© Chloe Stallion, Tamkeen Khan, Christian Guerrero, Stavros Tsipas, Gregory Wozniak. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 26.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Community-Acquired Pneumonia in Patients With Diabetes: Narrative Review

Yun Xie[1], MD, PhD; Ao Zhang[2], MCM; Ying Wang[3], MD, PhD; Ruilan Wang[1], MD, PhD

[1]Department of Critical Care Medicine, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Department of Critical Care Medicine, Shanghai General Hospital of Nanjing Medical University, Shanghai, China

[3]Shanghai Institute of Immunology, Department of Immunology and Microbiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiaotong University School of Medicine, Shanghai, China

Corresponding Author:
Ruilan Wang, MD, PhD
Department of Critical Care Medicine, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Abstract

Background: Patients with diabetes carry a 1.5- to 2-fold higher risk of community-acquired pneumonia (CAP) and experience more severe outcomes, yet the mechanisms that integrate metabolic dysregulation, pathogen shifts, and novel cell death pathways remain fragmented.

Objective: This study aimed to synthesize current evidence on epidemiology, pathophysiology, causative pathogens, clinical outcomes, and management of CAP in adults with diabetes and to identify research gaps for future trials.

Methods: A narrative review (1999 to August 2025) of PubMed, EMBASE, the Cochrane Library, and Web of Science was conducted. GRADE (Grading of Recommendations Assessment, Development, and Evaluation) was used to rate evidence from 81 selected English-language studies (randomized controlled trials, cohorts, and meta-analyses).

Results: Diabetes increases CAP incidence (relative risk 1.73, 95% CI 1.46 - 2.04), hospitalization (+30% - 50%), and 30-day mortality (odds ratio 1.67, 95 % CI 1.45-1.92). Key drivers include hyperglycemia-induced immune paralysis, pulmonary microangiopathy, ferroptosis, glycation and methylation changes, and gut-lung dysbiosis that collectively favor multidrug-resistant Gram-negative bacilli (*Klebsiella* and *Pseudomonas*) and severe viral and fungal coinfections. Host-targeted therapy with moderate glycemic control (5 - 10 mmol/L), continued metformin, and pathogen-directed antibiotics improves survival, whereas single-dose PCV20 and annual influenza vaccination prevents approximately 45% of CAP admissions. Emerging strategies (nanozymes, ferroptosis inhibitors, probiotics, and proteolysis-targeting chimeras) are still preclinical.

Conclusions: CAP in patients with diabetes is a distinct, more severe entity mediated by metabolic-immune crosstalk. Multicenter randomized controlled trials integrating tight glucose monitoring, novel host-directed agents, and microbiome modulation are warranted to translate mechanistic insights into better outcomes.

(JMIR Diabetes 2026;11:e82215) doi:10.2196/82215

KEYWORDS

diabetes mellitus; community-acquired pneumonia; pathophysiology; pathogens; complications; treatment; prevention; microbiome

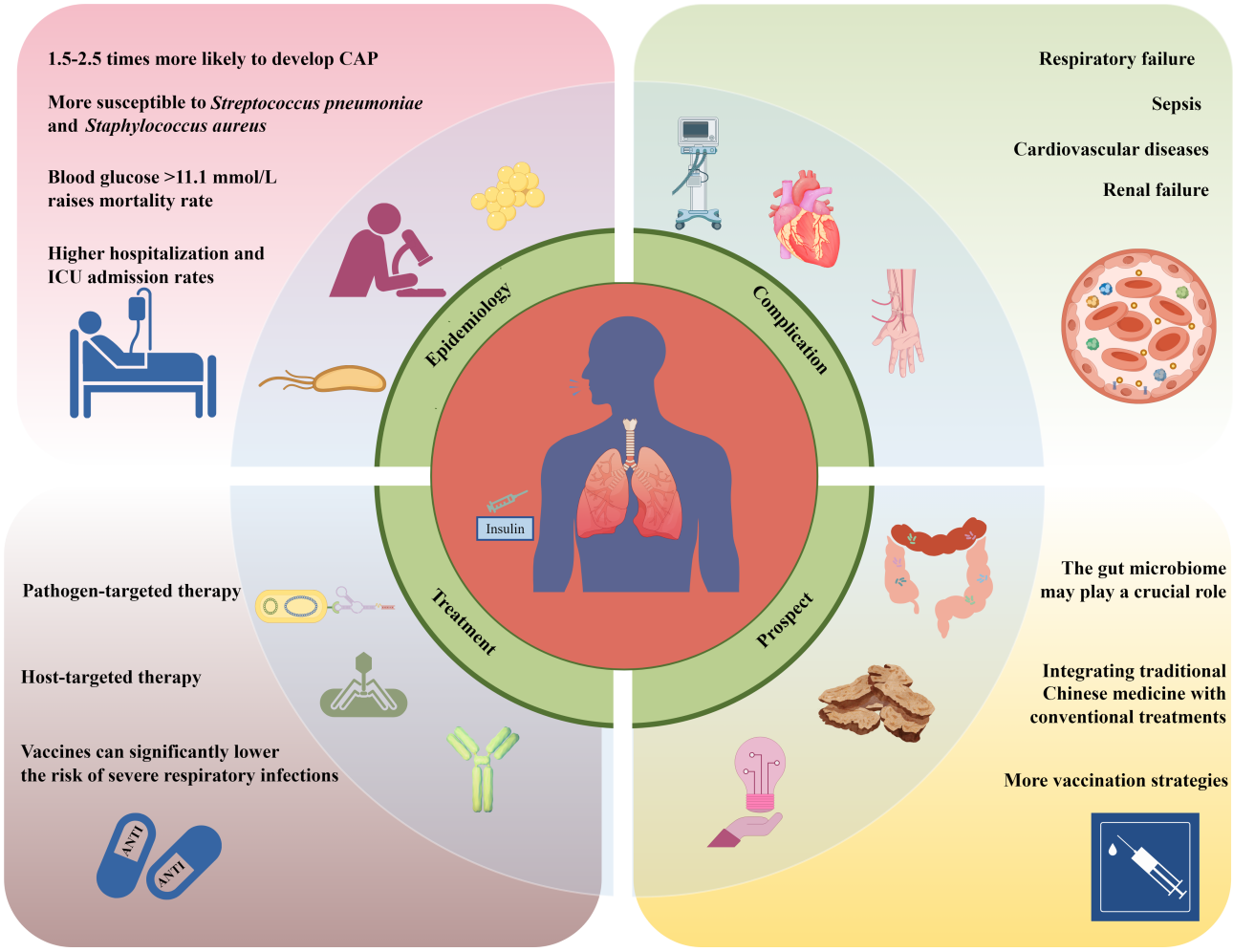
Introduction

Diabetes mellitus (DM), affecting more than 537 million individuals worldwide [1], significantly increases susceptibility to infections, including community-acquired pneumonia (CAP) [2]. CAP is a leading cause of morbidity and mortality worldwide, with patients with diabetes experiencing higher rates of hospitalization (30% - 50% higher), complications, and mortality (odds ratio [OR] 1.6) compared with those without diabetes [3,4]. The interplay of hyperglycemia, immune dysfunction, and comorbidities creates a “triple threat,” predisposing patients with diabetes to severe CAP and complicating treatment [5]. This review provides a comprehensive, critical analysis of CAP in patients with diabetes, integrating epidemiological data, pathophysiological mechanisms, pathogen profiles, clinical outcomes, and management strategies. It emphasizes novel research areas (eg, ferroptosis, microbiome influences, and epigenetic modifications) and evaluates the strength and limitations of current evidence, distinguishing between clinical, in vitro, and animal studies to clarify translational relevance (Figure 1).

Patients with diabetes have a 1.5- to 2.5-fold greater risk of pneumonia due to hyperglycemia and immune dysfunction [6]. They experience more severe infections and complications, leading to higher mortality rates. Targeted therapies and vaccines are crucial for managing pneumonia in patients with diabetes.

XSL•FO
RenderX

Figure 1. Multiple risks and influencing factors for community-acquired pneumonia (CAP) in patients with diabetes. ICU: intensive care unit.



Methods

This is a narrative review synthesizing evidence on CAP in patients with diabetes, focusing on epidemiology, pathophysiology, pathogens, outcomes, and management. We conducted a systematic literature search in PubMed, EMBASE, the Cochrane Library, and Web of Science from January 1999 to August 2025, using the following terms: (“diabetes mellitus” OR “type 2 diabetes”) AND (“community-acquired pneumonia” OR “CAP”) AND (epidemiology OR pathophysiology OR pathogens OR treatment OR prevention) (Multimedia Appendix 1). Inclusion criteria were English-language studies (randomized controlled trials [RCTs], cohorts, and meta-analyses) on adults aged 18 years or older; exclusion criteria were pediatric studies, hospital-acquired pneumonia, and nonhuman studies without translational relevance. We screened 1248 abstracts, reviewing 156 full texts; of these, 81% (156/1248) were included. Evidence was graded using GRADE (Grading of Recommendations Assessment, Development, and Evaluation): high (RCTs or meta-analyses), moderate (cohorts), low (case series or animal studies), and very low (in vitro studies or hypothesis). Quantitative claims include 95% CIs and designs where available.

Results

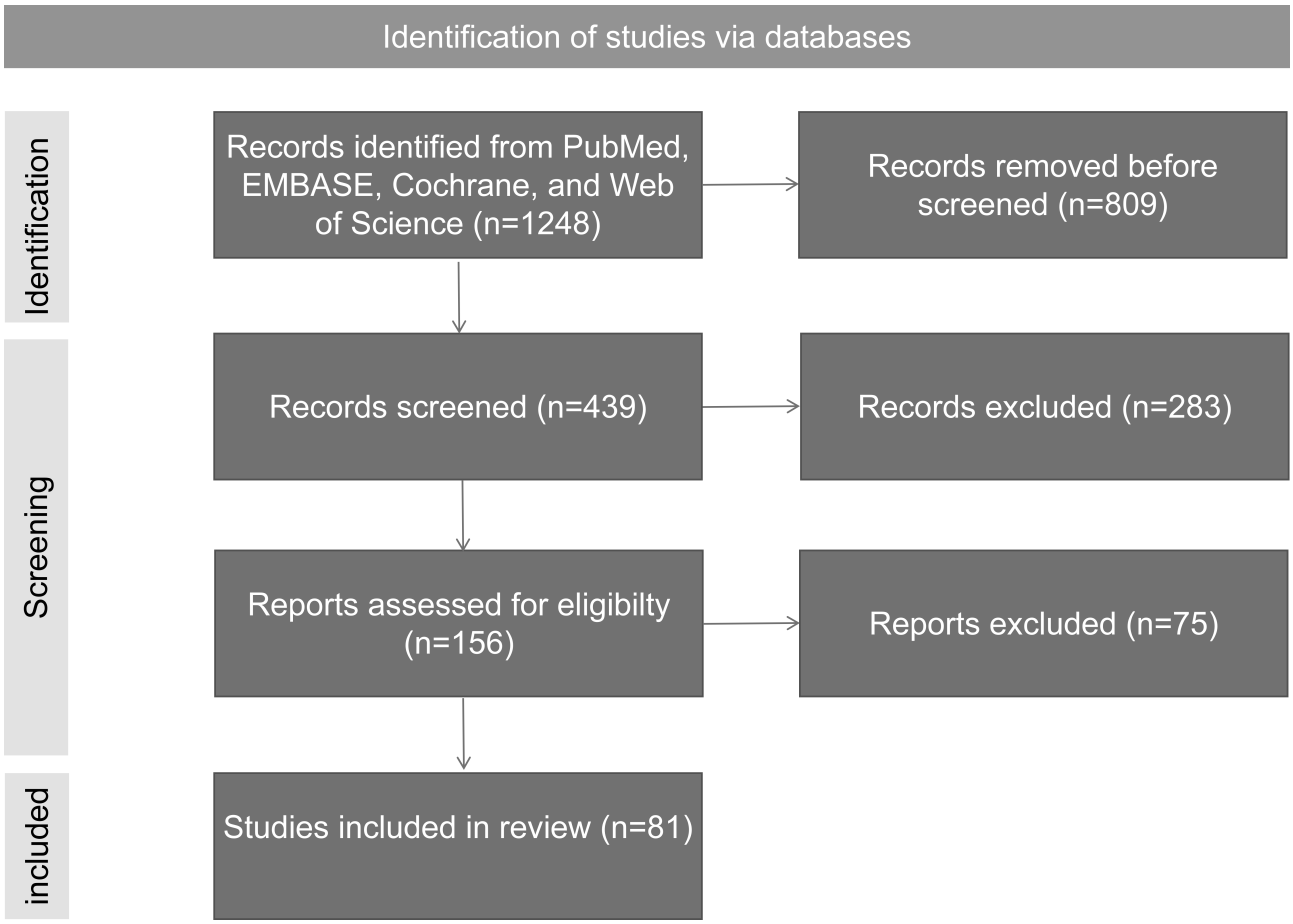
Epidemiology

The annual incidence of CAP ranges from 1.5 to 14 cases per 1000 adults, with patients with diabetes facing a 1.5- to 2-fold higher risk (10 - 20 cases per 1000) [7]. A 2024 meta-analysis of observational studies (pooled sample size >1 million) reported a relative risk of 1.73 (95% CI 1.46-2.04) for pneumonia in patients with diabetes, primarily those with type 2 diabetes, although the findings were limited by heterogeneity in metrics of glycemic control and the lack of RCTs [6]. Rates of hospitalization are 30% to 50% higher, and 30-day mortality odds are 1.6-fold greater in patients with diabetes and CAP, with observational studies showing consistent trends but with potential confounding by comorbidities [4,8]. Poor glycemic control (glycosylated hemoglobin A_{1c} [HbA_{1c}] ≥7%) doubles the risk of severe outcomes such as sepsis (odds ratio [OR] 2.0, 95% CI 1.4 - 2.8), with in-hospital mortality ranging from 10% to 20% in clinical cohorts [9]. Low- and middle-income countries (LMICs) face a higher CAP burden due to rising diabetes prevalence and limited health care access, although data are limited by underreporting [10]. People with DM were at higher risk for invasive pneumococcal disease (unadjusted OR 2.42, 95% CI 2.00-2.92) and case fatality rate (unadjusted OR 1.61, 95% CI 1.25-2.07), after adjusting for age, obesity, and smoking in propensity-matched cohorts (high,

XSL•FO
RenderX

meta-analyses) [4,6]. Diabetes significantly increases the probability of secondary pneumonia in patients with COVID-19 (moderate, time series [11]). Although observational studies consistently report a 1.5- to 2.0-fold risk, recent RCTs using continuous glucose monitoring (CGM) address confounding. In a 2025 RCT of 124 frail, critically ill patients with COVID-19 (30% with diabetes), CGM-guided glycemic control reduced 28-day intensive care unit (ICU) mortality by approximately 80% (hazard ratio 0.18, 95% CI 0.04 - 0.79), with benefit independent of baseline HbA_{1c} levels [12]. These findings underscore the need for more RCTs to confirm observational data and address biases such as selection and ascertainment (Figure 2, Table 1).

Figure 2. Flow diagram.



XSL•FO
RenderX

Table . Epidemiological features of community-acquired pneumonia (CAP) in patients with and without diabetes.

Outcome	Patients with diabetes	Patients without diabetes	Estimate (95% CI)	Study design and sample	Evidence grade[a]	Key references
Annual incidence	8 - 12/1000 adults	3 - 6/1000 adults	IRR[b] 1.66 (1.65 - 1.67)	National discharge cohort ≈900,000	High	Kornum et al [8] and López-de-Andrés et al [4]
RR[c] (any CAP)	—[d]	1 (ref)[e]	RR 1.26 (1.21 - 1.31)	Population-based case-control	High	Kornum et al [8]
CAP-related hospitalization	+26% - 66% versus non-DM[f]	Baseline	IRR 1.66 (1.65 - 1.67)	National cohort ≈900,000	High	López-de-Andrés et al [4]
30-d mortality after CAP	12% - 14%	13% - 14%	OR[g] 0.92 (0.91 - 0.94) lower	National cohort ≈900,000	High	López-de-Andrés et al [4]
Invasive pneumococcal disease	—	1 (ref)	OR 2.42 (2.00 - 2.92) unadjusted	Meta-analysis 36 studies >9 million	High	Silverii et al [6]
Case fatality rate (IPD[h])	—	1 (ref)	OR 1.61 (1.25 - 2.07) unadjusted	Meta-analysis 19 studies	High	Silverii et al [6]
Severe course or sepsis (HbA_{1c}[i] ≥7%)	—	1 (ref)	HR 1.52 (1.37 - 1.68)	National diabetes registry ≈500,000	High	Balintescu 2021 [9]

[a]Evidence grade: high=national registry or large meta-analysis; moderate=single prospective cohort; low=case series.
[b]IRR: incidence rate ratio.
[c]RR: relative risk.
[d]Not applicable.
[e]ref: reference.
[f]DM: diabetes mellitus.
[g]OR: odds ratio.
[h]IPD: invasive pneumococcal disease.
[i]HbA_{1c}: glycosylated hemoglobin A_{1c}.

Pathophysiological Mechanisms

Immune Dysregulation

Hyperglycemia impairs innate and adaptive immunity, reducing neutrophil chemotaxis, phagocytosis, and bactericidal activity, which are critical for clearing pathogens such as *Streptococcus pneumoniae* [13]. The nucleotide-binding oligomerization domain–like receptor thermal protein domain–associated protein 3 (NLRP3) inflammasome's hyperactivation drives excessive interleukin (IL)-1β and IL-6 production, exacerbating lung damage, as demonstrated in murine models of diabetes and pneumonia [14]. However, these models may not fully replicate human immune responses due to species-specific differences in inflammasome regulation. Decreased interferon-gamma production and altered T-cell responses, including reduced regulatory T cells, further compromise immunity, with in vitro studies showing a reduction in interferon-gamma in T cells from patients with diabetes [15]. A 2021 systematic review and meta-analysis (n=449,247) indicated that type 2 DM is associated with an approximately 2-fold increased risk of multidrug-resistant (MDR) bacterial infections, with underlying immune suppression in patients with diabetes proposed as a key contributing mechanism, although heterogeneity across included studies (I^2 up to 58.1%) should be noted [16].

Pulmonary Microangiopathy

Chronic hyperglycemia causes microvascular damage, impairing endothelial function and gas exchange, thereby facilitating pathogen colonization [17,18[19]]A retrospective cohort study demonstrated an increased risk of mortality in patients with severe CAP and type 2 diabetes complicated by microvascular disease, such as nephropathy and retinopathy, who often presented with computed tomography evidence of multilobar infiltrates [20]. These findings are supported by animal models showing increased vascular permeability in the lungs of animals with diabetes, although human studies are needed to validate these mechanisms.

Hyperglycemia and Glycemic Variability

Admission hyperglycemia (>11.1 mmol/L) is a strong predictor of poor CAP outcomes, with a 2024 meta-analysis (n=12 studies, >10,000 patients) reporting a pooled OR of 2.47 (95% CI 1.73 - 4.12) for short-term mortality [21]. Acute hyperglycemia exacerbates oxidative stress, while chronic hyperglycemia sustains endothelial dysfunction, as shown in clinical cohorts [22,23]. A 2015 retrospective cohort study (n=203) demonstrated that an elevated glycemic gap—a marker of acute glycemic variability—was associated with a 3 - to 4-fold increased risk of adverse outcomes, including acute respiratory failure requiring mechanical ventilation in patients with diabetes and CAP, underscoring the importance of dynamic glucose

XSL•FO
RenderX

monitoring [24]. Long-term CGM or serial HbA_{1c} trajectories are needed to determine whether sustained versus transient hyperglycemia drives pneumonia risk. Distinguishing acute versus chronic hyperglycemia effects remains a research hotspot, as most studies focus on admission glucose levels without longitudinal data.

Ferroptosis

Iron overload and hepcidin overexpression in patients with diabetes promote ferroptosis, an iron-dependent programmed cell death mechanism, exacerbating pulmonary inflammation and injury in murine models of diabetes (low, animal [25]). An in vitro study demonstrated that ferroptosis inhibitors (eg, ferrostatin-1) reduced lung epithelial cell death in high-glucose conditions, suggesting therapeutic potential [25]. In murine models of diabetes, ferroptosis was linked to acute respiratory distress syndrome (ARDS)–like pathology in CAP, although clinical trials are lacking to confirm these findings in humans. Emerging humanized models bridge the gap. In diabetes-derived organoids, high-glucose conditions upregulate NLRP3 and impair phagocytosis, recapitulating ferroptosis [26], human organoids show glutathione peroxidase 4 downregulation (very low, in vitro [26]). Targeting ferroptosis pathways could mitigate severe CAP outcomes, but challenges include identifying safe, specific inhibitors for more clinical use.

Glycation and Methylation

Nonenzymatic glycation alters protein structures (eg, angiotensin-converting enzyme 2), impairing function and worsening pneumonia severity, as shown in in vitro studies of lung tissue from patients with diabetes [27]. Methylation dysregulates immune gene expression, increasing CAP susceptibility, with a 2020 study identifying hypermethylation of IL-6 promoters in patients with diabetes and CAP [27]. Targeting glycation with inhibitors such as aminoguanidine or methylation with demethylating agents (eg, 5-azacytidine) shows promise in preclinical studies but requires RCTs to establish efficacy and safety.

RNA-Level Mechanisms

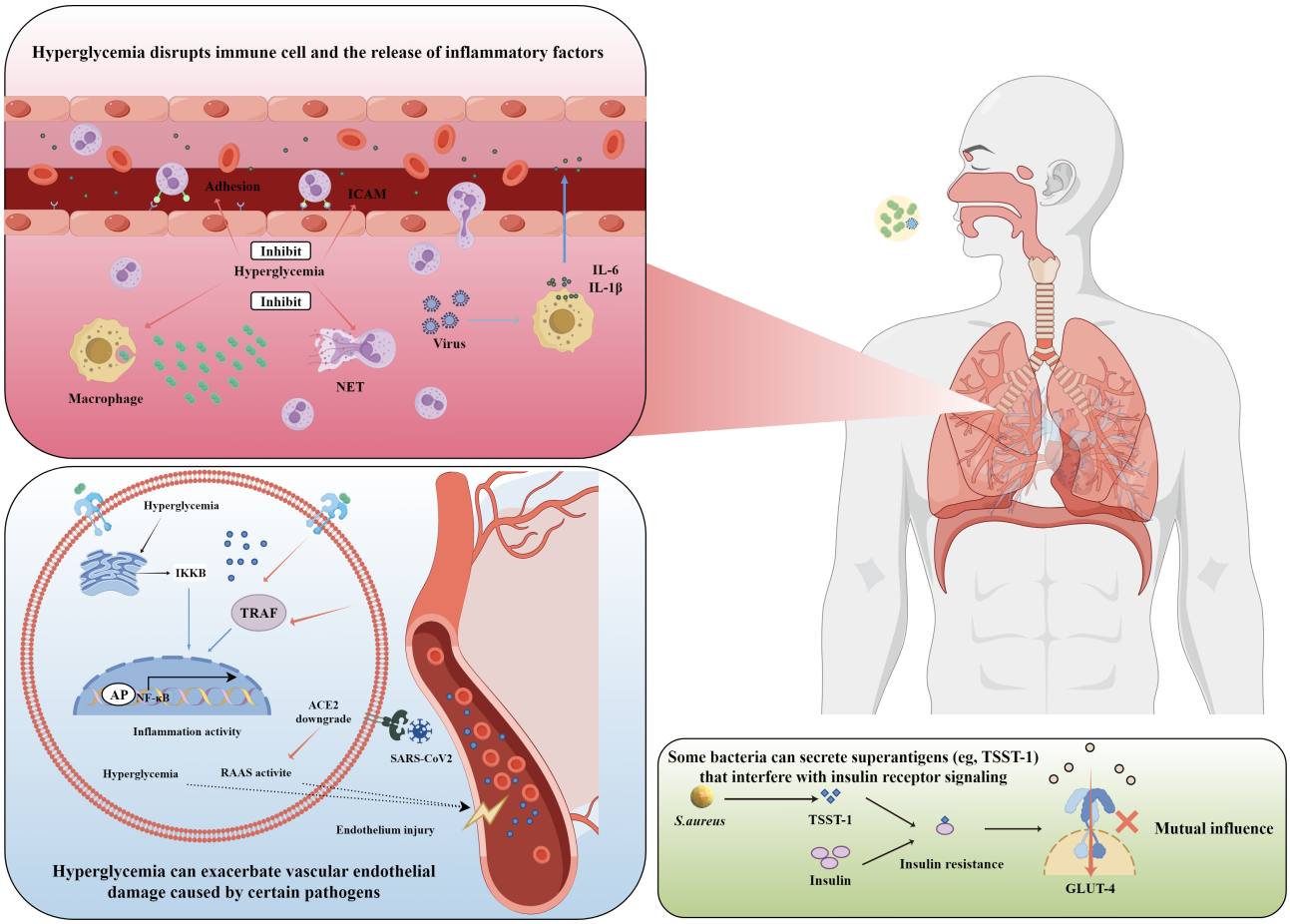
Gu et al [28] found that zinc finger E-box binding homeobox 1 antisense 1 (ZEB1-AS1) was downregulated in lung tissue from patients with diabetes and in high-glucose–treated BEAS-2B cells; this downregulation increased p53 and apoptosis in vitro. Hyperglycemia exacerbates this downregulation, increasing lung injury in murine models of diabetes. These findings suggest ZEB1-AS1 as a potential therapeutic target, but human studies are needed to validate its role and to develop RNA-based therapies, such as small interfering RNA or Clustered Regularly Interspaced Short Palindromic Repeats–based approaches.

Microbiome Alterations

Diabetes alters the lung and gut microbiome, promoting Gram-negative pathogens such as *Klebsiella pneumoniae* [29].[30] Zhou et al [31] conducted a multiomics longitudinal study of 86 participants with prediabetes or T2D risk, integrating 16S ribosomal RNA sequencing, metabolomics, and host cytokine profiles and found that insulin-resistant individuals exhibited *Akkermansia* depletion alongside elevated phenylalanine-associated metabolites, which correlated with increased respiratory infection events, supporting a gut-lung axis link. Fecal transplantation from diabetes-affected donors to mice increased *K pneumoniae* colonization, which was reversed by *Akkermansia* depletion. Preclinical studies in mice with diabetes suggest that probiotics (eg, *Lactobacillus*) could restore microbial balance and reduce CAP severity. On the basis of the pooled analysis of 9 randomized trials in a study by Manzanares et al [32], probiotics reduced the risk of new-onset ventilator-associated pneumonia by 26% (RR 0.74, 95% CI 0.61 - 0.90), although causality remains in the absence of CAP-specific RCTs. On the basis of the PROSPECT pilot trial (Cook et al [33]), 150 critically ill ventilated patients received *L rhamnosus* GG, achieving 97% protocol adherence and an observed ventilator-associated pneumonia rate of 19%, supporting feasibility for a larger RCT on probiotic prevention of ICU-acquired pneumonia. Additional clinical trials are needed to confirm efficacy (Figure 3, Table 2).

XSL•FO
RenderX

Figure 3. Pathophysiological mechanisms of community-acquired pneumonia in patients with diabetes. Hyperglycemia enhances macrophage inflammatory responses, releasing interleukin (IL)-6 and IL-1β, while impairing neutrophil chemotaxis and phagocytosis. Activation of nuclear factor kappa-B (NF-κB) and tumor necrosis factor receptor–associated factor (TRAF) pathways damages vascular endothelial cells, and ACE2 degradation exacerbates renin-angiotensin-aldosterone system (RAAS) activation, increasing cardiovascular complications. Bacterial superantigens (eg, toxic shock syndrome toxin-1 [TSST-1] from *S. aureus*) induce insulin resistance, creating a vicious cycle. ACE2: angiotensin-converting enzyme 2; AP: activator protein; GLUT-4: glucose transporter type 4; ICAM: intercellular adhesion molecule; IKKB: inhibitor of nuclear factor kappa B kinase subunit beta; NET: neutrophil extracellular trap.



XSL•FO
RenderX

Table . Pathophysiological mechanisms of community-acquired pneumonia (CAP) in patients with diabetes.

Mechanism	Description	Impact on CAP	References and evidence
Immune dysregulation	Persistent hyperglycemia impairs neutrophil chemotaxis, phagocytosis, and oxidative burst; NLRP3[b] inflammasome hyperactivation increases IL[c]-1β/IL-6, while IFN[d]-γ and regulatory T-cell responses are reduced.	Delayed clearance of *S. pneumoniae* and MDR[e] pathogens; ~2-fold higher risk of MDR infection; exaggerated lung inflammation.	Meta-analysis, n=449,247, I^2=58%, high quality [13-16]
Pulmonary microangiopathy	Chronic hyperglycemia thickens the alveolar-capillary basement membrane, increases microvascular permeability, and impairs gas exchange.	Promotes multilobar infiltrates; higher mortality when retinopathy and nephropathy coexist.	Retrospective cohort, moderate quality [17,18,20]
Hyperglycemia and glycemic variability	Admission glucose >11.1 mmol/L or wide glycemic excursions enhance oxidative stress and sustained endothelial NF-κB[f] activation.	Short-term mortality OR[g] 2.47 (95% CI 1.73 - 4.12); 3 - to 4-fold more likely to require mechanical ventilation.	Meta-analysis 12 studies, n >10,000, high quality [21-24]
Ferroptosis	Iron overload (↑ hepcidin, free Fe^{2+}) triggers iron-dependent lipid peroxidation and death of alveolar epithelial cells.	Amplifies lung injury; ARDS-like[h] histology in diabetic mice; ferrostatin-1 protective in vitro or rodent models.	Animal and organoid studies, low quality [25,26]
Glycation and methylation	Nonenzymatic glycation modifies ACE2[i], surfactant proteins; promoter hypermethylation maintains high IL-6 transcription.	Compromised protein function, sustained inflammatory gene expression, greater CAP severity.	Clinical+in vitro, low-moderate quality
RNA-level mechanisms	Hyperglycemia ↓ lncRNA ZEB1-AS1[j] → p53 ↑ → alveolar epithelial apoptosis.	Larger radiological lesions; siRNA[k] or CRISPR[l] modulation under investigation.	Human tissue+BEAS-2B cells, very low quality
Microbiome alterations	Gut-lung axis dysbiosis: *Akkermansia* depletion, expansion of *Klebsiella* and other G bacilli, accompanied by proinflammatory phenylalanine metabolites.	2 - to 3-fold rise in G or MDR CAP and recurrence; probiotics lower VAP[m] risk by 26% (RR[n] 0.74, 9 RCTs[o]), but CAP-specific RCTs pending.	Longitudinal multiomics+probiotic meta-analysis, moderate quality [29,31,32,34]

[b]NLRP3: NOD-like receptor thermal protein domain–associated protein 3.
[c]IL: interleukin.
[d]IFN: interferon.
[e]MDR: multidrug-resistant.
[f]NF-κB: nuclear factor kappa-B.
[g]OR: odds ratio.
[h]ARDS: acute respiratory distress syndrome.
[i]ACE2: angiotensin-converting enzyme 2.
[j] ZEB1-AS1: zinc finger E-box binding homeobox 1 antisense 1.
[k]siRNA: small interfering RNA.
[l]CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats.
[m]VAP: ventilator-associated pneumonia.
[n]RR: relative risk.
[o]RCT: randomized controlled trial.

Primary Pathogens

Bacterial Pathogens

Streptococcus pneumoniae accounts for approximately 8% of CAP cases in patients with diabetes, with higher bacteremic rates [4]. In the European multicenter CAPNETZ cohort (2002 - 2022), *Streptococcus pneumoniae* remained the predominant pathogen, identified in 36% of patients with diabetes and CAP and 39% of patients with CAP and without diabetes with an identified organism (2954/13,611), adults with diabetes and CAP exhibited a significantly higher isolation rate of *Enterobacteriaceae* (13% vs 8%; P<.005) and modest differences in Haemophilus influenzae (8%), atypical bacteria (9%), and viruses (22%) compared with individuals with diabetes, while *Staphylococcus aureus* and nonfermenting

XSL•FO
RenderX

Gram-negative bacilli each accounted for approximately 3% to 4% in both groups [35]. *Klebsiella pneumoniae* (15% - 20%) and *Pseudomonas aeruginosa* are prevalent, often multidrug-resistant, complicating treatment [36]. *Klebsiella pneumoniae* caused 13% of microbiologically confirmed CAP in European adults with diabetes (CAPNETZ 2002 - 2022) [35] and 78% of carbapenem-resistant isolates in a 2021 Indian cohort of patients with diabetes [37]; data from these settings indicate its importance in CAP among patients with diabetes, but population-based proportions in Asian LMICs remain to be established. A 2024 single-center prospective study from Mangalore, southern India, found that among hospitalized patients, the proportion of individuals with diabetes infected with carbapenem-resistant *Klebsiella pneumoniae* (CRKP) was approximately twice that of individuals without diabetes (78% vs 39%) [37]. *Legionella pneumophila* causes severe disease, especially in immunocompromised patients or patients with diabetes, and carries a higher mortality risk when diagnosis or appropriate antibiotic therapy is delayed [38]. In adult CAP across the Asia-Pacific region, *Streptococcus pneumoniae* remains the most frequently detected bacterium in patients with diabetes (≈10% - 24%), yet *K pneumoniae* (≈10% - 20%, up to 40% in Malaysia and Thailand), Haemophilus influenzae (≈5% - 24%), and *Pseudomonas aeruginosa* or *Burkholderia pseudomallei* (20% - 30% in ICU settings) are recovered far more often than reported in Western series. Atypical pathogens and mixed infections each account for roughly one-quarter of cases [39].

Viral Pathogens

Influenza and SARS-CoV-2 are significant, with patients with diabetes showing higher rates of severe pneumonia and ARDS [40]. Elevated ACE2 expression in patients with diabetes facilitates SARS-CoV-2 entry, worsening outcomes [41]. A 2023 study highlighted a 4.64-fold increased risk of COVID-19–related ARDS in patients with diabetes [42].

Fungal Pathogens

Opportunistic fungi such as *Aspergillus* and *Candida* are more common in patients with diabetes due to immune suppression, with pulmonary aspergillosis carrying a 50% to 70% mortality rate [43]. *Pneumocystis jirovecii* and mucormycosis are notable in patients with poorly controlled diabetes [44]. A 2020 study reported a higher incidence of invasive fungal infections in patients with diabetes, with the prevalence of invasive fungal disease among hospitalized patients with type 2 diabetes to be 0.4%, approximately double that of the 0.2% observed in inpatients without diabetes [44]. Fungal CAP remains uncommon (<5%), but post–COVID reports are rising; in patients with diabetes—a classic driver of mucormycosis—preguideline Indian series already showed attack rates of approximately 0.1 to 0.3 per 1000 and mortality of approximately 50% (low-certainty data [45]), a signal now echoed in Asian cohorts of patients with diabetes.

Atypical Pathogens

Mycoplasma salivarium and *Legionella* cause severe infections in patients with diabetes, often presenting atypically, necessitating early diagnosis [46]. A 2023 case report documented severe *M. salivarium* empyema in a patient with diabetes [46].

Clinical Outcomes

Patients with diabetes and CAP required invasive mechanical ventilation in 1.6 % of admissions (vs 2.1% in patients without diabetes) and had slightly higher overall risk-adjusted odds of in-hospital mortality, although the database did not capture ICU admission rates or radiographic extent of disease [4]. Hyperglycemia amplifies systemic inflammation and complement dysfunction, and population studies cited in 2024 show that diabetes carries an overall 1.5 to 4-fold increase in infection-related hospitalization, with the risk gradient most pronounced for sepsis, pneumonia, and renal infection [47]. In a 2023 Iranian cross-sectional study of 172 hospitalized patients with CAP, those with diabetes incurred a median CURB-65 score of 3 versus 2 in those without diabetes, required ICU admission more than 3 times as often (22% vs 7%), and stayed a mean of 8.5 days versus 7.9 days, underscoring diabetes as an independent predictor of more severe pneumonia course [48]. Complications include higher rates of pleural effusion and respiratory failure, with MDR pathogens increasing complication rates[49]. Among 600 patients with *K pneumoniae* infection, patients with diabetes exhibited 78% CRKP and 71% MDR rates versus 39% and 11% in those without diabetes and carried a 27% infection-related mortality (vs 2%), illustrating how MDR pathogens amplify severe outcomes in patients with diabetes [37] (Table 3).

XSL•FO
RenderX

Table . Primary pathogens causing community-acquired pneumonia (CAP) in adults with diabetes.

Pathogen	Proportion in DM[b]-CAP	Location/technique/setting	Mortality (95% CI)	Strength of evidence	Key references
Bacteria					
Streptococcus pneumoniae	8% - 36%	Global; sputum/BAL[c] culture+ PCR[d]; community and hospital	10% - 15% (8% - 18%)	High (multicenter cohort n≈13,000)	[4, 35]
Klebsiella pneumoniae (MDR[e]/CRKP[f])	10% - 20% (up to 40% ICU[g], Asia)	Asia-Pacific, India; culture; hospital and ICU	20% - 25% (15% - 30%)	Moderate (prospective center n≈600)	[36,37,39]
Haemophilus influenzae	5% - 24%	Asia-Pacific; culture/PCR; community	8% - 12%	Moderate (cohort n≈3000)	[35, 39]
Pseudomonas aeruginosa	3% - 4% (Western) → 20% - 30% (ICU, Asia)[h]	Culture; ICU and hospital	25% - 35%	Moderate (ICU cohorts)	[35, 39]
Staphylococcus aureus (including MRSA)[i]	3% - 4%	Culture; community and hospital	30% - 40%	Low (case series)	[35, 50]
Legionella pneumophila	1% - 3%	Urine Ag/PCR; community and travel associated	10% - 30%	Moderate (surveillance)	[38]
Viruses					
SARS-CoV-2	10% - 15% (COVID era)	PCR; community and hospital	ARDS[j] OR[k] 4.6 (3.2 - 6.7)	High (meta-analysis n>100 000)	[40,42]
Influenza A/B	5% - 10%	PCR; community and hospital	10% - 20%	High (seasonal surveillance)	[20]
Fungi					
Aspergillus spp. (including IPA[l])	<5% (0.4% invasive)	BAL GM[m]/culture/biopsy; ICU and immunosuppressed	50% - 70% (40% - 80%)	Low (case series)	[43,44]
Candida spp. (pulmonary)	<2%	BAL culture/biopsy; ICU	20% - 40%	Very low (case reports)	[43]
Mucorales (mucormycosis)	0.1 - 0.3% (diabetes cohort)	Biopsy/culture; ICU (post-COVID)	≈50%	Low (national surveys)	[44,45]
Atypical bacteria					
Mycoplasma pneumoniae	5% - 10%	PCR/serology; community	1% - 3%	Moderate (surveillance)	[39,46]
Legionella pneumophila[n]	1% - 3%	See above	—[o]	—	—

[b]DM: diabetes mellitus.

[c]BAL: bronchoalveolar lavage.

[d]PCR: polymerase chain reaction.

[e]MDR: multidrug-resistant.

[f]CRKP: carbapenem-resistant *Klebsiella pneumoniae.*

[g]ICU: intensive care unit.

[h]Proportions are median ranges from Western (CAPNETZ) and Asia-Pacific cohorts unless specified.

[i]MRSA: methicillin-resistant *Staphylococcus aureus*.

[j]ARDS: acute respiratory distress syndrome.

[k]OR: odds ratio.

[l]IPA: invasive pulmonary aspergillosis.

[m]GM: galactomannan.

[n]Listed once under “bacteria”; totals may overlap in mixed infections.

XSL•FO
RenderX

[o]Not applicable.

Treatment Strategies

Pathogen-Targeted Therapy

Bacterial Pneumonia

Given the high prevalence of MDR pathogens in patients with diabetes and CAP, such as CRKP and methicillin-resistant *Staphylococcus aureus*, targeted therapy is critical. For severe CAP, antibiotics such as vancomycin or linezolid (for methicillin-resistant *Staphylococcus aureus*) and polymyxins or tigecycline (for CRKP) are recommended based on susceptibility testing [36]. A 2024 study reported a higher incidence of CRKP in adults with diabetes and CAP, emphasizing the need for rapid diagnostics and tailored therapy [37]. Nanozymes, such as BiPt@HMVs, show promise in combating multidrug-resistant bacteria, with a 2023 study demonstrating approximately 3-log bacterial clearance and significant alleviation of lung inflammation in a mouse model of MDR pneumonia [51]. Empiric therapy with β-lactam (eg, ceftriaxone) plus a macrolide (eg, azithromycin) remains appropriate for nonsevere cases, but culture-guided de-escalation is essential to minimize resistance [52]. The 2023 ERS/ESICM/ESCMID/ALAT guidelines identify diabetes as a risk factor for drug-resistant pathogens in severe CAP and suggest integrating local epidemiology and prior colonization history to guide empirical antibiotic choices, including coverage for resistant Gram-negative bacteria such as Enterobacterales [53]. Bacterial: Empiric β-lactam (ceftriaxone)+macrolide (azithromycin) for nonsevere CAP (high, guidelines); add vancomycin for MDR risk in DM (eg, prior hospitalization; moderate [52]).

Viral Pneumonia

Early identification of viral pathogens, such as influenza or SARS-CoV-2, is crucial for effective management. Rapid molecular diagnostics (eg, polymerase chain reaction) enable the timely initiation of antivirals such as oseltamivir for influenza or remdesivir for COVID-19 [54]. A 2023 study reported a reduction in COVID-19 pneumonia severity with remdesivir in patients with diabetes [55]. Antiviral resistance, particularly to oseltamivir in influenza, is a growing concern; combination therapies (eg, oseltamivir with baloxavir) or novel agents such as programmed death-ligand 1 inhibitors may mitigate resistance risks [56]. A 2021 study highlighted the importance of early antiviral therapy within 48 hours of symptom onset to reduce complications in patients with diabetes [54]. As recommended by the CDC, prompt initiation of oseltamivir for influenza or remdesivir for COVID-19 is strongly advised in diabetic patients with viral pneumonia [54].

Fungal Pneumonia

Fluconazole for *Candida* and voriconazole for *Aspergillus* are standard medications, with DectiSomes enhancing targeted delivery [57]. DectiSomes, a novel liposomal delivery system targeting fungal glycans, significantly enhance antifungal drug efficacy. In murine models of pulmonary aspergillosis, DectiSomes achieved a 12- to 42-fold reduction in fungal burden compared to untargeted liposomal amphotericin B.

Host-Targeted Therapy

Glycemic Control

Maintaining blood glucose within a flexible range of 5 to 10 mmol/L is recommended to reduce mortality (OR 0.62) while minimizing the risk of hypoglycemia, which is a concern with stricter targets (4 - 8 mmol/L) in clinical practice [21]. Insulin protocols with frequent monitoring (eg, every 2 - 4 h) help stabilize glycemic variability, particularly in critically ill patients [22]. A 2022 study confirmed that moderate glycemic control (5 - 10 mmol/L) reduced ICU admission by 25% in patients with diabetes and CAP without significant hypoglycemic events [58]. Targets include 140 to 180 mg/dL (7.8 - 10.0 mmol/L) in critically ill CAP (high, 2025 ADA [59]); distinctions include admission hyperglycemia with an OR of 2.47 for mortality (95% CI 1.9 - 3.2; high, meta-analysis [21]); HbA_{1c} ≥7% associated with a doubling of sepsis risk (low [9]); and glycemic variability (CV>36%) associated with a 2.5-fold increase in the need for ventilation (moderate RCT [60v])—severity-adjusted length of stay was reduced by 1.7 days with glycemic control (moderate RCT, n=200 [48]).

Metformin

Metformin continuation was linked to 14% lower 30-day mortality in patients with diabetes and CAP (HR 0.86, 95% CI 0.78 - 0.95; moderate certainty, 2023 US Veterans cohort, n≈15,000 [61]). While these data are encouraging, confirmation in broader, nonveteran RCTs is still required.

Discussion

Adjunctive Therapies

Low-dose corticosteroids (eg, methylprednisolone 0.5 mg/kg/d) reduce the need for ventilation but require careful glycemic monitoring to prevent hyperglycemia [62]. Metformin continuation is linked to lower mortality. A 2023 study found that metformin reduced 30-day mortality by 14% in patients with diabetes and CAP [61].

Comorbidity Management

Optimizing heart failure and renal function reduces adverse events. Renal-dose antibiotic adjustments are critical [4]. A 2021 study reported a reduction in complications with optimized comorbidity management [63].

Management in LMICs

In LMICs, where DM prevalence is rising fastest, an estimated 537 million adults had DM in 2021, of whom 80.6% lived in LMICs [1]; scarce resources magnify the CAP burden [10]. Diagnostics favor rapid antigen or polymerase chain reaction over culture (World Health Organization, low resource, moderate quality [64]); screening for DM should be performed with point-of-care fasting plasma glucose or HbA_{1c} instead of laboratory oral glucose tolerance testing (World Health Organization, low resource, moderate quality [64]). Empirical therapy administers amoxicillin-clavulanate for nonsevere CAP (with adjustment for local resistance; Pan American Health Organization, high quality [64]) and deescalates using

XSL•FO
RenderX

biomarkers such as C-reactive protein. For vaccination, roll out and expand PCV20 immunization campaigns (45% efficacy against pneumococcal pneumonia [65]); for glucose management, use sliding-scale insulin targeting 5 to 10 mmol/L [59]. Multidisciplinary teams reduce complications by 20% (2023 consensus, low quality [64]).

Preventive Measures

Pneumococcal (PPSV23) and influenza vaccinations reduce CAP hospitalization by 30% in patients with diabetes, although efficacy may be reduced due to immune dysregulation [66]. A 2023 study demonstrated a 45.6% (95% CI 21.8% - 62.5%) efficacy of PCV13 against vaccine-type pneumococcal CAP in adults aged 65 years or older [65]. Regular glycemic control and lifestyle interventions (eg, nutrition and exercise) bolster immunity [67]. According to the 2024 Advisory Committee on Immunization Practices 2024 recommendations [68], pneumococcal vaccination with a single dose of PCV20 and PCV21 is recommended for adults aged 50 years or older with DM (high, guidelines; efficacy 45% - 60% vs vaccine-type CAP [65]); and adults aged 50 years or older with diabetes should receive timely pneumococcal vaccine to reduce the increased risk of CAP and invasive pneumococcal disease (high, guidelines [66]). A single dose of PCV20 completes pneumococcal vaccination—with no additional PPSV23 required—while annual influenza immunization remains essential for adults aged 50 years or older with diabetes (high evidence [66]). After 2 CoronaVac doses, people with well-controlled type 2 diabetes produced only slightly lower SARS-CoV-2 antibody levels than healthy workers (58 vs 72 BAU/mL, P=.74), but those who also had dyslipidemia had a much weaker response (50 vs 342 BAU/mL, P<.001) [69].

Future Research Directions

Autophagy Regulation

Autophagy-inducing compounds reduce SARS-CoV-2 propagation [70]. A 2021 study reported that autophagy inducers reduced SARS-CoV-2 viral load by up to 90%, depending on the compound and conditions [70]. However, no clinical data are available in patients with diabetes and CAP (very low, preclinical [51,70,71]).

Proteolysis-Targeting Chimeras

CYPA-ProTACs target host cofactors, enhancing antiviral effects [72]. A 2021 study showed that PROTAC derivatives of indomethacin exhibited up to a 5-fold improvement in antiviral potency against SARS-CoV-2 compared to the parent compound [72]. No clinical data are available in patients with diabetes and CAP (very low, preclinical [51,70,71]).

Nanoparticle Delivery

Phage-like nanocarriers improve antibiotic efficacy [73]. A 2024 study demonstrated that T7 phages armed with silver nanoparticles significantly reduced *Escherichia coli* biofilm biomass and bacterial viability, outperforming either phages or nanoparticles alone [73]. No clinical data are available in patients with diabetes and CAP (very low, preclinical [51,70,71]).

Microbiome Impact

Alterations in the gut and lung microbiome influence CAP susceptibility and severity in patients with diabetes [29]. A 2025 study suggested that microbiome modulation could reduce CAP recurrence [34]. Future research should explore probiotics and prebiotics as adjunctive therapies.

Vaccination Strategies

Evaluating the efficacy of adjuvanted or combination vaccines in patients with diabetes could enhance immunization outcomes [69]. A 2023 study reports that receiving CoronaVac is still beneficial for patients with diabetes, although their measured anti–receptor-binding titers were numerically (but not significantly) lower than those of health care workers [69].

Novel Therapies

Targeting ferroptosis, glycation, methylation, and lncRNA ZEB1-AS1 offers potential therapeutic avenues in patients with diabetes [25,27,28]. Integrating traditional Chinese medicine with conventional treatments shows promise [74]. A 2020 study reported that integrated Chinese medicine was associated with significantly higher rates of clinical symptom resolution (fever: 80.3% vs 53.1%, fatigue: 77.6% vs 53.8%, cough: 66.1% vs 42.9%, and sputum production: 85.3% vs 46.2%; P<.05 or P<.01) and a 13% relative reduction in hospitalization (10.98% vs 24.39%) compared with usual care alone [74].

Limitations and Future Directions

This review has several limitations. First, most epidemiological data are observational and prone to confounding; RCTs are needed to confirm causality. Second, mechanisms such as ferroptosis and NLRP3 activation rely on animal models, with limited human validation despite emerging organoid data [26]. Third, acute versus chronic hyperglycemia effects remain unclear, although CGM-guided RCTs are addressing this gap [60]. Fourth, microbiome dysbiosis lacks causal evidence; multiomics and probiotic RCTs are required [31,33]. Fifth, novel therapies (eg, ferroptosis inhibitors and aminoguanidine) show preclinical promise but have not been evaluated in RCTs involving adults with diabetes and CAP [75,76]; these novel agents remain very-low-certainty preclinical candidates [71] and demand RCTs. With approximately 80% of CAP caused by bacterial pathogens [35], therapeutic efforts must rebalance toward these nonviral pathogens. Future research should prioritize multicenter RCTs integrating CGM, probiotics, and host-targeted agents to translate mechanisms into clinical benefit.

Conclusions

Patients with diabetes face heightened CAP risks due to immune dysregulation, hyperglycemia, and microbial shifts, leading to severe disease, complications, and mortality. Early antibiotics targeting special pathogens, rapid antiviral therapy, and moderate glycemic control (5 - 10 mmol/L) are critical. Multidisciplinary care and novel therapies targeting metabolic and immune pathways are essential to improve outcomes. Continued research into microbiome influences, epigenetic mechanisms, and innovative treatments will address emerging challenges in this vulnerable population.

XSL•FO
RenderX

Funding

This study was supported by the Clinical Research Innovation Plan of Shanghai General Hospital (CCTR-2025C09), National Key Clinical Specialist Construction Project (Z155080000004), National Key Research and Development Program of China (2024YFC3044400), the Key Supporting Discipline of Shanghai Healthcare System (2023ZDFC0102), the Science and Technology of Shanghai Committee (23Y31900100), and Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0506502).

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

YX and RW were involved in the conception and design. All authors were involved in the writing of the manuscript. All authors were involved in the final approval of the manuscript.
YW is the co-corresponding author and can be reached via email at ywangssmu@shsmu.edu.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete electronic search strategies for PubMed, EMBASE, Cochrane CENTRAL, and Web of Science (1999-2025), showing Medical Subject Headings/Emtree terms, text words, and Boolean operators used to identify studies on diabetes and community-acquired pneumonia.
[DOCX File, 11 KB - diabetes_v11i1e82215_app1.docx]

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Cilloniz C, Torres A. Diabetes mellitus and pneumococcal pneumonia. Diagnostics (Basel) 2024 Apr 22;14(8):859. [doi: 10.3390/diagnostics14080859] [Medline: 38667504]
3. Ignatova GL, Blinova EV, Struch SV, Syrochkina MA. Risk of community acquired pneumonia in patients with diabetes mellitus: review. Ter Arkh 2022 Mar 15;94(3):448-453. [doi: 10.26442/00403660.2022.03.201447] [Medline: 36286912]
4. López-de-Andrés A, de Miguel-Díez J, Jiménez-Trujillo I, et al. Hospitalisation with community-acquired pneumonia among patients with type 2 diabetes: an observational population-based study in Spain from 2004 to 2013. BMJ Open 2017 Jan 5;7(1):e013097. [doi: 10.1136/bmjopen-2016-013097] [Medline: 28057653]
5. Darwitz BP, Genito CJ, Thurlow LR. Triple threat: how diabetes results in worsened bacterial infections. Infect Immun 2024 Sep 10;92(9):e0050923. [doi: 10.1128/iai.00509-23] [Medline: 38526063]
6. Silverii GA, Gabutti G, Tafuri S, et al. Diabetes as a risk factor for pneumococcal disease and severe related outcomes and efficacy/effectiveness of vaccination in diabetic population. Results from meta-analysis of observational studies. Acta Diabetol 2024 Aug;61(8):1029-1039. [doi: 10.1007/s00592-024-02282-5] [Medline: 38684540]
7. Torres A, Chalmers JD, Dela Cruz CS, et al. Challenges in severe community-acquired pneumonia: a point-of-view review. Intensive Care Med 2019 Feb;45(2):159-171. [doi: 10.1007/s00134-019-05519-y] [Medline: 30706119]
8. Kornum JB, Thomsen RW, Riis A, Lervang HH, Schønheyder HC, Sørensen HT. Diabetes, glycemic control, and risk of hospitalization with pneumonia: a population-based case-control study. Diabetes Care 2008 Aug;31(8):1541-1545. [doi: 10.2337/dc08-0138] [Medline: 18487479]
9. Balintescu A, Lind M, Franko MA, et al. Glycemic control and risk of sepsis and subsequent mortality in type 2 diabetes. Diabetes Care 2022 Jan 1;45(1):127-133. [doi: 10.2337/dc21-1050] [Medline: 34716212]
10. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
11. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481. [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
12. Shang J, Yuan Z, Zhang Z, Zhou Q, Zou Y, Wang W. Effectiveness of continuous glucose monitoring on short-term, in-hospital mortality among frail and critically ill patients with COVID-19: randomized controlled trial. J Med Internet Res 2025 Feb 7;27:e67012. [doi: 10.2196/67012] [Medline: 39918851]

XSL•FO
RenderX

13. Berbudi A, Rahmadika N, Tjahjadi AI, Ruslami R. Type 2 diabetes and its impact on the Immune system. Curr Diabetes Rev 2020 May 12;16(5):442-449. [doi: 10.2174/1573399815666191024085838]
14. Grailer JJ, Canning BA, Kalbitz M, et al. Critical role for the NLRP3 inflammasome during acute lung injury. J Immunol 2014 Jun 15;192(12):5974-5983. [doi: 10.4049/jimmunol.1400368] [Medline: 24795455]
15. Han M, Ma K, Wang X, et al. Immunological characteristics in type 2 diabetes mellitus among COVID-19 patients. Front Endocrinol (Lausanne) 2021;12:596518. [doi: 10.3389/fendo.2021.596518] [Medline: 33776910]
16. Carrillo-Larco RM, Anza-Ramírez C, Saal-Zapata G, et al. Type 2 diabetes mellitus and antibiotic-resistant infections: a systematic review and meta-analysis. J Epidemiol Community Health 2022 Jan;76(1):75-84. [doi: 10.1136/jech-2020-216029] [Medline: 34326183]
17. Knapp S. Diabetes and infection: is there a link?--A mini-review. Gerontology 2013;59(2):99-104. [doi: 10.1159/000345107] [Medline: 23182884]
18. Zhang X, Wu H, Fan B, et al. The role of age on the risk relationship between prediabetes and major morbidities and mortality: analysis of the Hong Kong diabetes surveillance database of 2 million Chinese adults. Lancet Reg Health West Pac 2023 Jan;30:100599. [doi: 10.1016/j.lanwpc.2022.100599] [Medline: 36419741]
19. Costantini E, Carlin M, Porta M, Brizzi MF. Type 2 diabetes mellitus and sepsis: state of the art, certainties and missing evidence. Acta Diabetol 2021 Sep;58(9):1139-1151. [doi: 10.1007/s00592-021-01728-4] [Medline: 33973089]
20. Yang H, Yue R, Zhou J, et al. Study on metabonomics of Chinese herbal medicine in the treatment of type 2 diabetes mellitus complicated with community-acquired pneumonia. Medicine (Baltimore) 2020;99(37):e22160. [doi: 10.1097/MD.0000000000022160]
21. Yuan S, Chen Y, Xie L. Association between glucose levels at admission and outcomes of pneumonia: a systematic review and meta-analysis. BMC Pulm Med 2024 Jul 30;24(1):369. [doi: 10.1186/s12890-024-03126-2] [Medline: 39080623]
22. Qambari H, Yu PK, Balaratnasingam C, Dickson J, Yu DY. Endothelial function can be modulated by acute hyperglycemia. Sci Rep 2025 Aug 20;15(1):30559. [doi: 10.1038/s41598-025-12612-4] [Medline: 40835626]
23. Zeng W, Huang X, Luo W, Chen M. Association of admission blood glucose level and clinical outcomes in elderly community‐acquired pneumonia patients with or without diabetes. Clinical Respiratory J 2022 Aug;16(8):562-571. [doi: 10.1111/crj.13526]
24. Chen PC, Liao WI, Wang YC, et al. An elevated glycemic gap is associated with adverse outcomes in diabetic patients with community-acquired pneumonia. Medicine (Baltimore) 2015 Aug;94(34):e1456. [doi: 10.1097/MD.0000000000001456] [Medline: 26313809]
25. Banchini F, Vallisa D, Maniscalco P, Capelli P. Iron overload and Hepcidin overexpression could play a key role in COVID infection, and may explain vulnerability in elderly, diabetics, and obese patients. Acta Biomed 2020 Sep 7;91(3):e2020013. [doi: 10.23750/abm.v91i3.9826] [Medline: 32921750]
26. Yang L, Han Y, Zhang T, et al. Human vascularized macrophage-islet organoids to model immune-mediated pancreatic β cell pyroptosis upon viral infection. Cell Stem Cell 2024 Nov 7;31(11):1612-1629. [doi: 10.1016/j.stem.2024.08.007] [Medline: 39232561]
27. Sartore G, Ragazzi E, Faccin L, Lapolla A. A role of glycation and methylation for SARS-CoV-2 infection in diabetes? Med Hypotheses 2020 Nov;144:110247. [doi: 10.1016/j.mehy.2020.110247] [Medline: 33254553]
28. Gu L, Sun H, Yan Z. lncRNA ZEB1-AS1 is downregulated in diabetic lung and regulates lung cell apoptosis. Exp Ther Med 2020 Dec;20(6):225. [doi: 10.3892/etm.2020.9355] [Medline: 33193839]
29. Fliegerová KO, Mahayri TM, Sechovcová H, et al. Diabetes and gut microbiome. Front Microbiol 2024;15:1451054. [doi: 10.3389/fmicb.2024.1451054] [Medline: 39839113]
30. Yang J, Li J, Zhang L, et al. Highly diverse sputum microbiota correlates with the disease severity in patients with community-acquired pneumonia: a longitudinal cohort study. Respir Res 2024 May 29;25(1):223. [doi: 10.1186/s12931-024-02821-2] [Medline: 38811936]
31. Zhou X, Shen X, Johnson JS, et al. Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. Cell Host Microbe 2024 Apr 10;32(4):506-526. [doi: 10.1016/j.chom.2024.02.012] [Medline: 38479397]
32. Manzanares W, Lemieux M, Langlois PL, Wischmeyer PE. Probiotic and synbiotic therapy in critical illness: a systematic review and meta-analysis. Crit Care 2016 Aug 19;19(1):262. [doi: 10.1186/s13054-016-1434-y] [Medline: 27538711]
33. Johnstone J, Meade M, Marshall J, et al. Probiotics: prevention of severe pneumonia and endotracheal colonization trial—PROSPECT: protocol for a feasibility randomized pilot trial. Pilot Feasibility Stud 2015 Dec;1(1):19. [doi: 10.1186/s40814-015-0013-3]
34. Wikki I, Palmu J, Kauko A, et al. Prospective association between the gut microbiota and incident pneumonia: a cohort study of 6419 individuals. Respir Res 2025 Dec 22;26(1):354. [doi: 10.1186/s12931-025-03453-w] [Medline: 41430301]
35. Millet Pascual-Leone B, Fiocca Vernengo F, Hillus D, et al. Community-acquired pneumonia in diabetic patients is characterised by a distinct pathogen spectrum and enhanced inflammation: results from CAPNETZ, a prospective observational cohort study. Infection 2025 Oct 12. [doi: 10.1007/s15010-025-02659-w] [Medline: 41077620]
36. Aon M, Aoun AH, Al Shami A, et al. Association of diabetes mellitus with increased mortality in carbapenem-resistant enterobacterales infections. Cureus 2024 Feb;16(2):e53606. [doi: 10.7759/cureus.53606] [Medline: 38449962]

XSL•FO
RenderX

37. Pattolath A, Adhikari P, Pai V. Carbapenemase-producing Klebsiella pneumoniae infections in diabetic and nondiabetic hospitalized patients. Cureus 2024 Jan;16(1):e52468. [doi: 10.7759/cureus.52468] [Medline: 38371145]
38. Cunha BA, Burillo A, Bouza E. Legionnaires' disease. Lancet 2016 Jan 23;387(10016):376-385. [doi: 10.1016/S0140-6736(15)60078-2] [Medline: 26231463]
39. Song JH, Thamlikitkul V, Hsueh PR. Clinical and economic burden of community-acquired pneumonia amongst adults in the Asia-Pacific region. Int J Antimicrob Agents 2011 Aug;38(2):108-117. [doi: 10.1016/j.ijantimicag.2011.02.017] [Medline: 21683553]
40. Zhu B, Wei X, Narasimhan H, et al. Inhibition of the mitochondrial pyruvate carrier simultaneously mitigates hyperinflammation and hyperglycemia in COVID-19. Sci Immunol 2023 Apr 14;8(82):eadf0348. [doi: 10.1126/sciimmunol.adf0348] [Medline: 36821695]
41. Kass-Gergi S, Zhao G, Wong J, et al. Disruption of immune responses by type 1 diabetes exacerbates SARS-CoV-2 mediated lung injury. Am J Physiol Lung Cell Mol Physiol 2024 Dec 1;327(6):L839-L851. [doi: 10.1152/ajplung.00250.2024] [Medline: 39320093]
42. Garg A, Posa MK, Kumar A. Diabetes and deaths of COVID-19 patients: systematic review of meta-analyses. Health Sci Rev (Oxf) 2023 Jun;7:100099. [doi: 10.1016/j.hsr.2023.100099] [Medline: 37229298]
43. Klekotka RB, Mizgała E, Król W. The etiology of lower respiratory tract infections in people with diabetes. Pneumonol Alergol Pol 2015;83(5):401-408. [doi: 10.5603/PiAP.2015.0065] [Medline: 26379004]
44. Lao M, Li C, Li J, Chen D, Ding M, Gong Y. Opportunistic invasive fungal disease in patients with type 2 diabetes mellitus from Southern China: clinical features and associated factors. J Diabetes Investig 2020 May;11(3):731-744. [doi: 10.1111/jdi.13183] [Medline: 31758642]
45. Cornely OA, Alastruey-Izquierdo A, Arenz D, et al. Global guideline for the diagnosis and management of mucormycosis: an initiative of the European Confederation of Medical Mycology in cooperation with the Mycoses Study Group Education and Research Consortium. Lancet Infect Dis 2019 Dec;19(12):e405-e421. [doi: 10.1016/S1473-3099(19)30312-3] [Medline: 31699664]
46. Weng NC, Yu WL. Isolation of Mycoplasma salivarium in the empyema of a diabetic patient with deep neck infection and necrotizing mediastinitis: a case report. IDCases 2023;34:e01915. [doi: 10.1016/j.idcr.2023.e01915] [Medline: 37886697]
47. Holt RIG, Cockram CS, Ma RCW, Luk AOY. Diabetes and infection: review of the epidemiology, mechanisms and principles of treatment. Diabetologia 2024 Jul;67(7):1168-1180. [doi: 10.1007/s00125-024-06102-x] [Medline: 38374451]
48. Hashemi SH, Saki F, Borzouei S, Bawand R, Soltanian A. Comprehensive comparison of clinicoradiological, laboratory, and prognostic factors of community-acquired pneumonia in diabetic and nondiabetic hospitalized patients. Turk J Med Sci 2023;53(6):1776-1785. [doi: 10.55730/1300-0144.5747] [Medline: 38813518]
49. Jensen AV, Baunbæk Egelund G, Bang Andersen S, et al. The glycemic gap and 90-day mortality in community-acquired pneumonia. A prospective cohort study. Ann Am Thorac Soc 2019 Dec;16(12):1518-1526. [doi: 10.1513/AnnalsATS.201901-007OC] [Medline: 31437014]
50. Vardakas KZ, Matthaiou DK, Falagas ME. Incidence, characteristics and outcomes of patients with severe community acquired-MRSA pneumonia. Eur Respir J 2009 Nov;34(5):1148-1158. [doi: 10.1183/09031936.00041009] [Medline: 19541719]
51. Yao H, Zhou R, Wang J, et al. Pathogen‐targeting bimetallic nanozymes as ultrasonic‐augmented ROS generator against multidrug resistant bacterial infection. Adv Healthcare Mater 2023 Oct;12(25):2300449. [doi: 10.1002/adhm.202300449]
52. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med 2019 Oct 1;200(7):e45-e67. [doi: 10.1164/rccm.201908-1581ST] [Medline: 31573350]
53. Martin-Loeches I, Torres A, Nagavci B, et al. ERS/ESICM/ESCMID/ALAT guidelines for the management of severe community-acquired pneumonia. Eur Respir J 2023 Apr;61(4):2200735. [doi: 10.1183/13993003.00735-2022] [Medline: 37012080]
54. Jeon MH, Chung JW, Choi SH, Kim TH, Lee EJ, Choo EJ. Pneumonia risk factors and clinical features of hospitalized patients older than 15 years with pandemic influenza A (H1N1) in South Korea: a multicenter study. Diagn Microbiol Infect Dis 2011 Jun;70(2):230-235. [doi: 10.1016/j.diagmicrobio.2011.01.005] [Medline: 21596224]
55. Amstutz A, Speich B, Mentré F, et al. Effects of remdesivir in patients hospitalised with COVID-19: a systematic review and individual patient data meta-analysis of randomised controlled trials. Lancet Respir Med 2023 May;11(5):453-464. [doi: 10.1016/S2213-2600(22)00528-8] [Medline: 36828006]
56. Ning H, Chiu SH, Xu X, Ma Y, Chen JL, Yang G. The immunosuppressive roles of PD-L1 during influenza a virus infection. Int J Mol Sci 2023 May 11;24(10):8586. [doi: 10.3390/ijms24108586] [Medline: 37239931]
57. Meagher RB, Lewis ZA, Ambati S, Lin X. Aiming for a bull's-eye: targeting antifungals to fungi with dectin-decorated liposomes. PLoS Pathog 2021 Jul;17(7):e1009699. [doi: 10.1371/journal.ppat.1009699] [Medline: 34293050]
58. Barmanray RD, Cheuk N, Fourlanos S, Greenberg PB, Colman PG, Worth LJ. In-hospital hyperglycemia but not diabetes mellitus alone is associated with increased in-hospital mortality in community-acquired pneumonia (CAP): a systematic review and meta-analysis of observational studies prior to COVID-19. BMJ Open Diabetes Res Care 2022 Jul;10(4):e002880. [doi: 10.1136/bmjdrc-2022-002880] [Medline: 35790320]

XSL•FO
RenderX

59. American Diabetes Association Professional Practice Committee. 6. Glycemic goals and hypoglycemia: standards of care in diabetes—2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S128-S145. [doi: 10.2337/dc25-S006] [Medline: 39651981]
60. Shang J, Yuan Z, Zhang Z, Zhou Q, Zou Y, Wang W. Effectiveness of Continuous Glucose Monitoring on Short-Term, In-Hospital Mortality Among Frail and Critically Ill Patients With COVID-19: Randomized Controlled Trial. J Med Internet Res 2025 Feb 7;27:e67012. [doi: 10.2196/67012] [Medline: 39918851]
61. Mohammed T, Bowe M, Plant A, Perez M, Alvarez CA, Mortensen EM. Metformin use is associated with lower mortality in veterans with diabetes hospitalized with pneumonia. Clin Infect Dis 2023 Apr 3;76(7):1237-1244. [doi: 10.1093/cid/ciac900] [Medline: 36575139]
62. Hong S, Wang H, Li S, Liu J, Qiao L. A systematic review and meta-analysis of glucocorticoids treatment in severe COVID-19: methylprednisolone versus dexamethasone. BMC Infect Dis 2023 May 5;23(1):290. [doi: 10.1186/s12879-023-08280-2] [Medline: 37147596]
63. Badora-Musiał K, Sagan A, Domagała A, Kowalska-Bobko I. Testing the 2017 PHC reform through pilots: strengthening prevention and chronic care coordination. Health Policy 2021 Feb;125(2):185-190. [doi: 10.1016/j.healthpol.2020.10.014] [Medline: 33298319]
64. Atun R, Davies JI, Gale EAM, et al. Diabetes in sub-Saharan Africa: from clinical care to health policy. Lancet Diabetes Endocrinol 2017 Aug;5(8):622-667. [doi: 10.1016/S2213-8587(17)30181-X] [Medline: 28688818]
65. Ghit A. Pneumococcal vaccination in diabetic patients: review from clinical practice. Egypt J Intern Med 2023;35(1):16. [doi: 10.1186/s43162-023-00202-7]
66. Torres A, Blasi F, Dartois N, Akova M. Which individuals are at increased risk of pneumococcal disease and why? Impact of COPD, asthma, smoking, diabetes, and/or chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. Thorax 2015 Oct;70(10):984-989. [doi: 10.1136/thoraxjnl-2015-206780] [Medline: 26219979]
67. Bader MS, Yi Y, Abouchehade K, Haroon B, Bishop LD, Hawboldt J. Community-acquired pneumonia in patients with diabetes mellitus: predictors of complications and length of hospital stay. Am J Med Sci 2016 Jul;352(1):30-35. [doi: 10.1016/j.amjms.2016.02.032] [Medline: 27432032]
68. Pneumococcal vaccine recommendations. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/pneumococcal/hcp/vaccine-recommendations/index.html [accessed 2026-01-22]
69. Dechates B, Porntharukchareon T, Sirisreetreerux S, et al. Immune response to CoronaVac and its safety in patients with type 2 diabetes compared with healthcare workers. Vaccines (Basel) 2023 Mar 17;11(3):684. [doi: 10.3390/vaccines11030684] [Medline: 36992267]
70. Gassen NC, Papies J, Bajaj T, et al. SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. Nat Commun 2021 Jun 21;12(1):3818. [doi: 10.1038/s41467-021-24007-w] [Medline: 34155207]
71. Yen FS, Wei JCC, Shih YH, Hsu CC, Hwu CM. Metformin use and the risk of bacterial pneumonia in patients with type 2 diabetes. Sci Rep 2022 Feb 28;12(1):3270. [doi: 10.1038/s41598-022-07294-1] [Medline: 35228620]
72. Desantis J, Mercorelli B, Celegato M, et al. Indomethacin-based PROTACs as pan-coronavirus antiviral agents. Eur J Med Chem 2021 Dec 15;226:113814. [doi: 10.1016/j.ejmech.2021.113814] [Medline: 34534839]
73. Szymczak M, Pankowski JA, Kwiatek A, et al. An effective antibiofilm strategy based on bacteriophages armed with silver nanoparticles. Sci Rep 2024 Apr 20;14(1):9088. [doi: 10.1038/s41598-024-59866-y] [Medline: 38643290]
74. Duan C, Xia WG, Zheng CJ, et al. Clinical observation on Jinhua Qinggan granule combined with conventional western medicine therapy in treating mild cases of coronavirus disease 2019 [Article in Chinese]. J Tradit Chin Med 2020;61(17):1473-1477. [doi: 10.13288/j.11-2166/r.2020.17.001]
75. Yang XD, Yang YY. Ferroptosis as a novel therapeutic target for diabetes and its complications. Front Endocrinol 2022;13. [doi: 10.3389/fendo.2022.853822]
76. Freedman BI, Wuerth JP, Cartwright K, et al. Design and baseline characteristics for the aminoguanidine Clinical Trial in Overt Type 2 Diabetic Nephropathy (ACTION II). Control Clin Trials 1999 Oct;20(5):493-510. [doi: 10.1016/s0197-2456(99)00024-0] [Medline: 10503809]

Abbreviations

ARDS: acute respiratory distress syndrome
CAP: community-acquired pneumonia
CGM: continuous glucose monitoring
CRKP: carbapenem-resistant *Klebsiella pneumoniae*
DM: diabetes mellitus
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HbA_{1c}: glycosylated hemoglobin A_{1c}
ICU: intensive care unit
IL: interleukin
LMIC: low- and middle-income country
MDR: multidrug-resistant

XSL•FO
RenderX

NLRP3: NOD-like receptor thermal protein domain–associated protein 3
OR: odds ratio
RCT: randomized controlled trial

Edited by S Li; submitted 11.Aug.2025; peer-reviewed by A Batta, M Ekwueme, SCL Au; revised version received 02.Jan.2026; accepted 02.Jan.2026; published 10.Mar.2026.

Please cite as:
Xie Y, Zhang A, Wang Y, Wang R
Community-Acquired Pneumonia in Patients With Diabetes: Narrative Review
JMIR Diabetes 2026;11:e82215
URL: https://diabetes.jmir.org/2026/1/e82215
doi:10.2196/82215

© Yun Xie, Ao Zhang, Ying Wang, Ruilan Wang. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 10.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Evolving Landscape of Continuous Glucose Monitoring Metrics in Type 1 Diabetes: Narrative Literature Review

Eric Williams Jr[1], MSHA, MSHI; Casey Rand[1], MSDS; Alessandra Ayers[2], BA; Cindy Ho[2], BA; Ashley Dunova[2], BA; David Klonoff[3], MD; Juan Espinoza[1], MD

[1]Stanley Manne Children's Research Institute, Lurie Children's Hospital, 225 East Chicago Avenue, Chicago, IL, United States

[2]Diabetes Technology Society, Burlingame, CA, United States

[3]Diabetes Research Institute, Mills-Peninsula, San Mateo, CA, United States

Corresponding Author:
Eric Williams Jr, MSHA, MSHI
Stanley Manne Children's Research Institute, Lurie Children's Hospital, 225 East Chicago Avenue, Chicago, IL, United States

Abstract

Background: Continuous glucose monitoring (CGM) has transformed diabetes management and research by providing high-frequency data that address many of the limitations of hemoglobin A_{1c}, enabling more precise clinical treatment targets and responsive trial endpoints. The richness and complexity of high-resolution time-series CGM data have spurred the development of numerous metrics for both clinical care and research applications. Beyond established metrics, there is a growing set of clinical, composite, and research-oriented measures that may support clinical decision support, intervention planning, risk stratification, and discovery-oriented research. This proliferation has created significant challenges in metric selection, interpretation, calculation, and standardization, particularly when metrics are applied across different devices, populations, software packages, and study designs.

Objective: The objective is to map the current landscape of CGM metrics and address ongoing challenges in metric selection, clinical and research interpretation, and standardization. We further sought to distinguish between metrics primarily suited for routine clinical interpretation and those designed to explore more granular or multidimensional features of glycemia in research settings.

Methods: We identified the literature focusing on the calculation, application, and interpretation of the following categories of CGM metrics: (1) standardized, (2) clinical, (3) emerging, and (4) composite. CGM metrics included in this study were identified from the 27 metrics included in the Diabetes Research Hub platform, additional published standardized and composite metrics, metrics used in established CGM analysis software, and emerging metrics identified during review. We narratively reviewed each metric's definitions, calculation methods, interpretation, clinical and research utility, and strengths and limitations. In total, 102 articles were reviewed, supporting the synthesis of 36 distinct CGM-derived metrics.

Results: The review identifies a fundamental divide in the CGM metric landscape. Standardized and clinical metrics, including time in range, mean glucose, coefficient of variation, and similar, prioritize simplicity and actionability. These metrics facilitate rapid decision-making in clinical settings but potentially mask granular glycemic fluctuations, event patterns, and discordance between average glucose values and variability. Emerging and composite metrics offer deeper insights into glycemic patterns, risk, and variability. However, many rely on specialized software or complex formulas, lack standardized thresholds or clear relationships to clinical outcomes, and do not have consensus methods of calculation and interpretation, limiting their adoption and hindering cross-study comparison.

Conclusions: While consensus exists for core clinical metrics, the lack of standardization for complex metrics hinders research replicability and clinical translation. Bridging this gap requires moving toward consensus metric definitions, open-science frameworks, and standardized code libraries. Metric selection should be guided by intended use. Clinical metrics should be well-established, interpretable, and actionable. Research metrics should be clearly described, reproducible, and linked to meaningful outcomes. This review provides a comprehensive resource for navigating the diverse spectrum of CGM metrics, clarifying their applications and limitations to support both research and clinical investigation.

(JMIR Diabetes 2026;11:e92455) doi:10.2196/92455

KEYWORDS

continuous glucose monitoring; CGM; time-series glucose data; glycemic variability; hemoglobin A1c; diabetes mellitus; diabetes research; narrative review

XSL•FO
RenderX

Introduction

Diabetes mellitus affects over 500 million people worldwide [1]. Maintaining optimal glycemic control remains critical for preventing the devastating complications that characterize this disease. For decades, hemoglobin A_{1c} (HbA_{1c}) has served as the gold standard biomarker for assessing long-term glycemic control, with extensive evidence linking elevated HbA_{1c} levels to increased risk of diabetes-related complications [2]. However, HbA_{1c} is a highly summarized, aggregate measure. It reflects a 2 - 3 month average and lacks information about the daily glycemic excursions that patients experience [2]. HbA_{1c} cannot capture the frequency or severity of hypoglycemic episodes, postprandial excursions, or glycemic variability, all factors that impact patient safety, quality of life, and clinical decision-making.

Continuous glucose monitoring (CGM) has emerged as a transformative technology that overcomes these fundamental limitations [3-6]. CGM provides real-time, high-frequency glucose measurements through subcutaneously inserted sensors, offering an unprecedented, high-resolution view of an individual's glycemic profile. Unlike single-point measurements, CGM systems can capture hundreds to thousands of measures daily, providing a more complete glycemic picture and revealing previously invisible patterns and fluctuations. This comprehensive view of glucose dynamics has enabled more personalized diabetes management and has fundamentally changed clinical care paradigms. Usage of CGM metrics in clinical practice has led to improved glycemic control, reduced hypoglycemia, and enhanced quality of life for people with diabetes [5,7,8]. Recently, evidence has emerged supporting the use of CGM metrics for monitoring in early-stage type 1 diabetes, and to predict disease progression and severity [9,10]. Consequently, CGM has rapidly become a standard component of diabetes care across type 1 and type 2 diabetes [11]. Beyond its clinical utility, CGM has transformed diabetes research by providing investigators with a powerful tool to characterize the glycemic effects of lifestyle and behavioral factors, identify predictors of glucose variability, and rigorously evaluate emerging therapeutic interventions [12-15].

Despite the potential of using CGM data for both clinical care and research, the volume and complexity of data generated by CGM technology present substantial analytical challenges. Thousands of time-series CGM readings per patient must be transformed into standardized, interpretable, and clinically meaningful summary metrics. To meet this challenge, numerous novel CGM-based metrics have been developed, including time-based metrics (eg, time in range [TIR]), glycemic summary statistics (eg, mean glucose and coefficient of variation [CV]), and measures to characterize excursions (eg, hyperglycemic episodes), among others. Existing CGM metrics include both well-established and widely accepted summary measures, often prioritizing simplicity and interpretability, and used to guide clinical management, as well as an extensive suite of emerging and research metrics designed to provide further insight. While consensus expert recommendations identifying core CGM metrics for clinical practice and research have been published [16-19], the quickly evolving landscape of metric development and usage in contemporary clinical research remains incompletely characterized. Specifically, there is a limited consolidated understanding of which metrics are being prioritized in clinical care and research, how these metrics are defined and calculated across different studies, and what specific outcomes they are intended to measure. These gaps lead to challenges in the interpretation and replication of CGM literature, limiting progress toward establishing standardized analytical frameworks that could accelerate both research innovation and clinical implementation [20].

Therefore, the purpose of this review is to examine recent literature to provide a comprehensive assessment of the current state of CGM metrics in diabetes care and research. Specifically, this review aims to (1) identify the full spectrum of CGM-derived metrics commonly being used; (2) document the calculation methods, threshold definitions, and applications of these metrics; (3) summarize their reported strengths, limitations, and interpretations; (4) offer clarity for researchers designing CGM protocols and analytic plans; (5) offer clarity for clinicians interpreting CGM-based evidence; and (6) facilitate the translation of complex CGM data into actionable clinical insights that improve diabetes knowledge, care, and patient outcomes.

Methods

Study Design

This study was conducted as a narrative review to map the evolving landscape of CGM metrics. A narrative review was selected because the main objective of this study is to identify existing CGM metrics, document their calculation methods, usability, strengths, and limitations to work toward standardization.

Literature Search

A search of PubMed (search date: November 14, 2025) and Google Scholar was conducted to identify relevant literature. To facilitate reproducibility, the search included 2 strategies. First, we searched for the phrase, continuous glucose monitoring metrics. Next, we paired each specific CGM metric with that same phrase (eg, "[metric name] continuous glucose monitoring metrics"); the following CGM metrics were included in the search: (1) the names of each of 27 CGM metrics included as part of the Diabetes Research Hub (DRH) [21] (Tables 1 and 2), a leading open-science repository for CGM data sharing and research; (2) additional metrics identified in previously published CGM review articles [22-27]; (3) metrics calculated by well-established CGM analysis software packages [25-28]; and (4) through input from diabetes clinicians and researchers involved in this project.

XSL•FO
RenderX

Table . Standardized and clinical metrics.

Metric name	Description	Range or formula	Clinical utility
TIR[a] [16,18,29-36]	Percentage of time glucose is 70 - 180 mg/dL	% of readings and time 70 - 180 mg/dL (3.9 - 10 mmol/L)	Quantifies time spent within standard glycemic range.
TITR[g] [16,18,37-39]	Stricter version of TIR, using 70 - 140 mg/dL	% of readings and time 70 - 140 mg/dL (3.9 - 7.8 mmol/L)	Used to assess tighter glycemic control in specific populations.
TAR H[h] [18,40]	Time spent in mild hyperglycemia	% of readings and time 181 - 250 mg/dL (10.1 - 13.9 mmol/L)	Measures time spent above target range but below critical threshold.
TAR VH[i] [18,40]	Time spent in severe hyperglycemia	% of reading and time >250 mg/dL (13.9 mmol/L)	Identifies periods of very elevated glucose levels.
TBR L[j] [18,40]	Time spent in mild hypoglycemia	% of readings and time 54 - 69 mg/dL (3.0 - 3.8 mmol/L)	Categorizes low glucose events that are not immediately severe.
TBR VL[k] [18,40]	Time spent in severe hypoglycemia	% of readings and time <54 mg/dL (3.0 mmol/L)	Indicates severe hypoglycemic episodes requiring intervention.
Mean glucose [29,41]	Average glucose over time	∑Glucose measurementsNumber of measurements	Estimates overall glycemic control from CGM[b] data.
Glucose variability (%CV) [4,18,36,37]	Degree of glucose fluctuation	Standard DeviationMean Glucose×100	Measures fluctuations in glucose levels over time.
GMI[c] [30,31]	Estimate HbA_{1c}[d] from CGM data	GMI (%)=3.31+0.02392 [mean glucose in mg/dL] GMI (mmol/mol)=12.71+4.70587 [mean glucose in mmol/L]	Converts mean glucose to estimated HbA_{1c} over the life of the CGM sensor.
AUC[e] [22,42,43]	Indicator of glycemic excursions in a specified period	AUC=∑i=1n−1(Gi+Gi+12)(ti+1−ti)	Helps compare interventions by assessing glycemic excursion and risk over time.
Percent wear time [18,44]	% of time CGM is worn	(Total Time CGM ActiveTotal Time Period)×100	Indicates completeness of CGM data collection.
Days of wear [18,44]	Number of days CGM was worn	% Active Wear Time×Total Monitoring Period (days)100	Reflects duration of CGM usage for data reliability.
eA1c [31]	Tracking of average glycemia and the estimation of HbA_{1c}	eA1c(t)=0.9512 * eA1c (t-1 d)+0.0488 * f (SMBG)	Fills the gap between infrequent laboratory HbA_{1c} tests. Highly predictive of the Hemoglobin Glycation Index (HGI).
Hypoglycemic episodes [17,45-48]	Event with glucose <70 mg/dL for ≥15 min	NHypo=∑i=1nI(Gi<70)	Defined by glucose thresholds and duration criteria.
Euglycemic episodes [35,49-51]	Glucose within normal range	NEugly=∑i=1nI(70≤Gi≤180)	Identifies periods of normoglycemia. May include atypical conditions like EDKA[f]
Hyperglycemic episodes [18,52-55]	Glucose >180 mg/dL for ≥15 min	NHyper=∑i=1nI(Gi>180)	Categorizes elevated glucose events based on severity.

[a]TIR: time in range.
[b]CGM: continuous glucose monitoring.
[c]GMI: glucose management indicator.
[d]HbA_{1c}: hemoglobin A_{1c}.
[e]AUC: area under the curve.
[f]EDKA: euglycemic diabetic ketoacidosis.
[g]TITR: time in tight range
[h]TAR H: time above range high
[i]TAR VH: time above range very high
[j]TBR L: time below range low
[k]TBR VL: time below range very low

XSL•FO
RenderX

Table . Emerging and composite metrics.

Metric name	Description	Range or formula	Clinical utility
MAGE[a] [36]	The "gold standard" for assessing short-term within-day glycemic variability.	MAGE = ∑λχ if λ>v λ=blood glucose changes from peak to nadir =number of valid observations V=1 SD mean glucose for 24-h period	It is used as a measure of glycemic lability, though the LI[b] may correlate more closely with clinical assessment of lability.
CONGA[c] [37,38,56]	A novel method developed to assess intraday glycemic variability.	CONGA=∑t=−t1t1(Dt−D¯)2k−1 D=∑t=t1tkDtK Dt=Gt−Gt−m k=number of observations with an observation n×60 min ago m=n×60 G=glucose measured	Distinguishes between patterns of glycemic control.
Mean of daily differences [12,37,42,57]	Measure of day-to-day variability.	k=number of observations with an observation 24 h ago G=glucose measured t=time (in min)	Determines if the short-term CGM[d] trace is representative of the 3-month glycemic pattern.
LI [4,41]	A measure of glycemic lability.	LI=∑n=1N−1(Gt−Gn+1)2(tn+1−tn) k=number of observations with an observation 24 h ago n=total number of readings in a week t=time	Used to quantify labile glucose control. It correlates more closely than MAGE with the clinical assessment of lability.
GCI[e] [42]	A novel measure developed to quantify the shape of glycemic variability.	Calculate and extract log-periodogram summary. Calculate the GCI by deriving the weights from the summary. Apply canonical correlation analysis to the weights to produce a raw GCI score. GCI∝∑i=16wimi	Used for the analysis of the shape of glycemic variability.
GRI[f] [43,58]	A single-number summary that assesses the quality of glycemia from CGM tracings.	Hypoglycemia component =VLow+(0.8×Low) Hyperglycemia Component =VHigh+(0.5×High) GRI = (3.0×VLow) + (2.4×Low) + (1.6×VHigh) + (0.8×High)	Assesses glycemic quality in clinical practice. Its primary benefit is to quickly evaluate glycemic control.
M value [34,37]	An index used to describe glucose variability and the quality of glycemic control.	M=∑t=t1tk|10×log Gt×18IGV|3N G=glucose measured IGV=ideal glucose values K=total number of observations N=total number of readings	Risk marker used to assess the quality of glycemic control.
J Index [24,46,59,60]	A universal index proposed for the assessment of current glucose control in diabetic patients.	J=0.001(MBG + SD))^2 for glucose measured in mg/dL J=0.324(MBG + SD)^2 for glucose measured in mmol/L	Highly sensitive to both the glycemic level and glycemic variations.
GRADE[g] [22,37,61]	An integrated risk score that summarizes the clinical risks of both hypoglycemia and hyperglycemia.	mmol/L GRADE value =425×[log (log (X))+0.16]2 mg/dL GRADE value =425×[log (log (X×18))+0.16]2 Where X=blood glucose	Used as an adjunct to HbA_{1c}[h] to report the degree of risk associated with glycemic variability.
ADRR[i] [33,62]	A measure used to assess blood glucose variability.	ADRR =1N∑t=1N[LR+HR] N=total number of readings LR=risk value attributed to low glucose HR=risk value attributed to high glucose	Adequate indicator of hyperglycemia.
LBGI[j] [29]	Quantifies the frequency and extent of low blood glucose readings to provide an optimal prediction of hypoglycemia risk.	LBGI =1n∑i=1nrl(χi)	Predicts the risk of severe hypoglycemia and helps split overall glucose variation into hypo- and hyperglycemia risk components.

XSL•FO
RenderX

Metric name	Description	Range or formula	Clinical utility
HBGI[k] [4,52,56,57]	Quantifies the frequency and extent of high blood glucose readings to predict the risk of hyperglycemia.	HBGI =1n∑i=1nrl(χi)	Predicts the risk of hyperglycemia, correlates with HbA_{1c}, and is associated with hyperglycemic excursions.
IGC[l] [23,24]	Composite measure of overall glycemic control.	IGC =Hypoglycemia Index + Hyperglycemia Index	Provides a comprehensive metric that is highly sensitive for detecting improvements in glycemic control following therapeutic interventions.
HAS[m] [22,59]	Combines hypoglycemia incidence and HbA_{1c} change.	Change in A_{1c}, the rate of SH[n] events, and a change in the rate of SH events.A score of HAS≥75 is arbitrarily assigned to represent a good to excellent intervention.	Allows for easier comparison of qualitatively different diabetes interventions using a single score.
Qscore [22,60]	A composite metric used to evaluate CGM profiles and screen for profiles needing therapeutic action.	Q-Score =8MG+7mg+2G+3G2+G3+7MD+10 Very Good:<4.0; Good: 4.0 - 5.9; Satisfactory: 6.0 - 8.4; Fair: 8.5 - 11.9; Poor:≥12.0	Stratifies metabolic control (eg,<4.0 is “very good”; ≥12.0 is “poor”); identifies individual CGM parameters that require optimization for patient-tailored therapy.
PGS[o] [22]	Combines TIR[p], mean glucose, Glycemic Variability Percentage (GVP), and the frequency and severity of hypoglycemia episodes.	PGS =f(GVP)+ g(MG)+ h(PTIR)+ j(N54,N70)	Identify problem areas in glycemic control to guide modification of therapy and complements A_{1C} assessment.
The Hypo-Triad [22,61]	Integrates hypoglycemic AUC[q], duration/time, and frequency of episodes to quantify the severity and risk of hypoglycemia.	IntHypo=|AUCx-Time|=AUC2+Time2	IntHypo describes the immediate clinical impact, while HypoRV represents the patient’s overall hypoglycemic environment, providing supplementary insights.
Glucose Pentagon and CGP[r] [22,62]	Provides a visual pentagonal shape and numerical score by integrating five glucose-centric CGM variables.	The area (ACGP) is calculated by summing the areas of the five resulting 72° triangles: ACGP =AToR CV+ ACV-IntHYPO+ AIntHYPO-IntHYPER+ AIntHYPER Mean glucose + AMean glucose-ToR, where:Ax-y = ½ a×b×sin γ (γ=angle between the axes, each 72°;a, b=length of each axis; A=area of triangle; and,x-y=length of axis for each area).	Enables visual and numerical comparison of glycemic control against an idealized non-diabetic state.
COGI[s] [22]	Evaluates the quality of control using TIR, TBR[t], and SD.	Index ranges between 0 and 100TIR (50% weight), TBR <70 mg/dL (35% weight), and SD (15% weight). %CV[u] may be used instead of SD	Provides a focused assessment on the 3 components considered most crucial for achieving safe glycemic control.

XSL•FO
RenderX

Metric name	Description	Range or formula	Clinical utility
Acute Glycemic Gap (GMI-HbA_{1c}) [46]	An independent risk factor for longer inpatient stays.	GMI-HbA_{1c}	Captures the discrepancy between patients' CGM readings and their HbA_{1C}.

[a]MAGE: mean average glucose excursions.
[b]LI: lability index.
[c]CONGA: continuous overall net glycemic action.
[d]CGM: continuous glucose monitoring.
[e]GCI: glucose color index.
[f]GRI: glycemic risk index.
[g]GRADE: glycemic risk assessment diabetes equation.
[h]HbA_{1c}: hemoglobin A_{1c}.
[i]ADRR: average daily risk range.
[j]LGBI: low blood glucose index.
[k]HGBI: high blood glucose index.
[l]IGC: index of glycemic control.
[m]HAS: hypoglycemia score.
[n]SH: severe hypoglycemia.
[o]PGS: personal glycemic state.
[p]TIR: time in range.
[q]AUC: area under the curve.
[r]CGP: comprehensive glucose profile.
[s]COGI: continuous glucose monitoring index.
[t]TBR: time below range.
[u]%CV: glucose variability.

Article Screening

A standardized screening process was developed to assess the relevance of retrieved articles. Screening information for each article, including the article objective, CGM-derived metrics discussed, and details regarding metric calculation, handling, and relevance, was recorded in Airtable (Formagrid Inc), a cloud-based relational database platform used to organize, link, and track project records. Moreover, 3 team members reviewed the titles, abstracts, and criteria questions to determine each article's relevance to the aim of the study. Discrepancies in inclusion were resolved through consensus or consultation with an additional lead researcher (refer to Authors' Contributions section). Articles were included for full-text review if they reported on the use, handling, or calculation of CGM-derived metrics. The results of the data extraction process are detailed in the results section.

Results

Literature Search and Article Screening

The search and screening process resulted in 102 articles flagged for full review, with article dates spanning from 1995 to 2025. Table 3 describes the standardized fields captured in the data extraction process. The included articles provided the foundational data for key details highlighting the calculation, relevance, definitions, and functionality of 36 CGM metrics. Data on each metric were summarized across articles by expert review.

XSL•FO
RenderX

Table . Data extraction fields captured for all included articles.

Focus and data field	Description
Article	
Study objective	The main research objective or primary endpoint of the study.
Prioritized metrics	The CGM[a] metrics highlighted as primary outcomes or key findings.
Relevance	A summary of the article's direct relevance to the review's objectives.
Technology	The CGM devices and software tools used for data collection and analysis.
Metric	
Definition	Purpose and detailed definition of the metric.
Range and interpretation	Published ranges of the metric and the interpretation of these values.
Formula	The specific formula or methodology used to calculate the metric.
Author recommendations	Author recommendations regarding the use, interpretation, strengths, and limitations of the metric.

[a]CGM: continuous glucose monitoring.

Overview of Identified CGM Metrics

Our review of the literature identified a total of 36 distinct CGM-derived metrics currently used for either diabetes clinical care, research, or both. Based on their application, calculation complexity, and degree of adoption, these metrics were categorized into two primary groups: (1) standard and clinical metrics and (2) emerging and composite metrics.

Standardized and Clinical Metrics

Overview

The standardized and clinical metrics (Table 1) comprise a core set of measures that have achieved widespread consensus and adoption. These metrics form the foundation of modern diabetes management, prioritizing simplicity, actionability, and interpretability for real-world diabetes management. They are designed to be easily understood by both health care professionals and people with diabetes, optimized for rapid interpretation during patient visits to enable real-time clinical decision-making. Overall, these metrics align well with established diabetes management practices and can be readily incorporated into existing clinical workflows and electronic health records. Standardized and clinical CGM metrics can be broadly categorized into several distinct categories described below, each serving specific purposes in diabetes care.

Composite Visual Reports

The ambulatory glucose profile (AGP) provides a standardized, single-page visual report of CGM data [18,22], summarizing key metrics, including composite indices like the glucose management indicator (GMI), facilitating the clinical interpretation of overall glycemic quality [17,58]. Daily glucose profiles are now typically presented in the AGP for several days [12,63] to aid in identifying typical 24-hour glucose patterns across treatment periods [18,63].

Time-Based Glycemic Metrics

These metrics quantify the percentage of time spent within defined glucose thresholds [18], and are widely used in clinical care [18,45]. Key metrics include TIR, time below range TBR; low <70 mg/dL and very low <54 mg/dL, and time above range TAR; high >180 mg/dL and very high >250 mg/dL [18]. A primary limitation is the inherent bias toward hyperglycemia due to the asymmetrical distribution of blood glucose values, which gives disproportionately heavier weight to high excursions in statistical calculations [29]. Consequently, metrics impacted by both low and high measures of glycemia, such as TIR, may lack sensitivity to changes in low values, such as longitudinal improvements in TBR. This limitation highlights the inadequacy of time-based measures in capturing the full clinical risk associated with hypo- or hyperglycemia as compared with composite indices [58].

Summary Statistical Metrics

These metrics characterize overall and long-term glycemic exposure, reflecting average glucose and stability over periods typically ranging from weeks to months [23]. These measurements include mean glucose, glucose variability (commonly %CV), and the estimated average measure GMI. This category also includes critical data sufficiency metrics such as percent wear time and days of wear. A significant limitation of summary metrics, like mean glucose and GMI, is that they do not reflect glucose fluctuations or glycemic variability [22,29,45,63], meaning significantly varied glucose profiles can yield identical mean values [63]. GMI often shows clinically significant discordance with measured HbA_{1c} (≥0.5%, 6 mmol/mol) due to nonglycemic factors influencing red blood cell turnover, reinforcing caution against interpreting GMI as interchangeable with HbA_{1c} [18,30,31]. Studies introduced an updated calculation of the GMI, known as the updated glucose management indicator (uGMI). The uGMI is derived from a kinetic model, and it is the expected HbA_{1c} under a steady glucose level [32]. Furthermore, estimates of glycemic variability often have a larger relative error compared with estimates of mean glucose [23] due to their sensitivity to noise and extreme values.

Event-Based Metrics

These metrics focus on discrete episodes of acute glycemic extremes, providing actionable data crucial for assessing patient

XSL•FO
RenderX

safety and guiding immediate therapeutic adjustments [17,18,33,63]. Unlike time-based metrics, these metrics include quantifying dangerous excursions such as hypoglycemic events (levels 1 and 2, eg, <70 mg/dL and <54 mg/dL) and hyperglycemic events (levels 1 and 2, eg, >180 mg/dL and >250 mg/dL) [18], alongside critical metabolic incidents (eg, euglycemic diabetic ketoacidosis). A primary limitation of event-based metrics is that CGM accuracy is often lower in the hypoglycemic range [17,18,34]. Also, the precise calculation of the number of events requires strict adherence to standardized definitions regarding the specific glucose thresholds and the required minimum duration (≥15 min) that must be maintained below or above the threshold [45], and subsequent events must be separated by a return to the target range for a specified time (eg, ≥15 min above 70 mg/dL) [17]. Without these standards, definitions can lead to over- or underestimation of event incidence [17]. For diagnosing acute events during periods of apparent euglycemia, such as euglycemic diabetic ketoacidosis, relying solely on glucose monitoring is insufficient, necessitating separate monitoring of metabolic acidosis and ketones [35]. Evaluating the frequency of hypoglycemic and hyperglycemic events is often considered more clinically meaningful than solely reporting the percentage of time spent in the respective ranges [45].

Emerging and Composite Metrics in Research Settings

While standard clinical CGM metrics are widely used in both research and clinical practice, emerging and composite metrics are gaining interest due to their potential to capture more complex aspects of glycemic control. Given the complexity of these metrics, variability in definitions, and evolving clinical relevance, it is challenging to compare results across different studies. These metrics often require specialized knowledge or software for calculation and interpretation, which limits their practical application in day-to-day clinical care. Furthermore, they may not have clear, immediate implications for treatment adjustments in routine clinical care. Metrics such as glycemic risk index (GRI) and glycemic risk assessment diabetes equation (GRADE) may eventually transition to clinical use as validation grows, but currently remain secondary to broadly endorsed measures. This review organizes the emerging and composite metrics in three categories: (1) metrics of glycemic variability, (2) composite indices of glycemic quality and risk, and (3) discordance metrics. This section provides a more detailed overview to aid interpretation and highlight their potential utility in future research and clinical decision-making. Table 2 summarizes each emerging and composite metric included in this review.

Metrics of Glycemic Variability

Mean Amplitude of Glycemic Excursions

Mean average glucose excursions (MAGE) focuses on the minimum to maximum span of blood glucose levels and is widely considered the gold standard for assessing short-term glycemic variability [36]. It was developed as a measure to characterize glycemic variability specifically for postprandial glucose excursions above the mean glucose level. MAGE calculates the mean of glycemic excursions by identifying nadir to peak blood glucose levels. The amplitude from peak to nadir glucose averages that exceed 1 SD of glucose are included in the final MAGE calculation.

Previously, researchers evaluated MAGE at multiple 2-day periods within a 14-day CGM sensor lifespan, finding these equivalent to MAGE calculated from longer (3 - 4 day) periods [36]. The study suggested that short-term periods are sufficient. MAGE lacks standardization in its calculation. There is no standard defining whether the amplitude should be calculated using ascending limbs, descending limbs, or a weighted average of both limbs. It requires specialized software programs for accurate calculations, limiting the metric's ability to be used in real-time practice or integrated into electronic health records.

Continuous Overall Net Glycemic Action

The continuous overall net glycemic action (CONGA) was originally developed for data from continuous interstitial glucose monitoring [37,38], and can be described as an index of the amount of time spent in glycemic excursions and the degree of glycemic variation that an individual experiences [37]. CONGA is particularly useful for helping clinicians and patients understand the magnitude and timing of blood glucose fluctuations in order to understand and optimize the patient's degree of glycemic control. The normative reference range for CONGA (mean +/− 2 SD) for healthy patients is 3.6 - 5.5, but may vary slightly by different ethnicities [39]. CONGA includes all data points, which means it may fail to effectively separate meaningful glycemic excursions from data noise [37].

Mean of Daily Differences

The mean of daily differences (MODD) was originally developed as a measure of intraday glycemic variability [12]. MODD is calculated using pairs of blood glucose measurements from CGM that are 24 hours apart. MODD was originally introduced as a metric to be used to better understand the magnitude of glycemic variability through CGM data. MODD is highly correlated with other measures of variability, such as CONGA and SD. The average normative reference range for MODD for healthy patients is 0.0 - 3.5 (SD 2), varying slightly for different ethnicities [39].

Lability Index

The lability index (LI) is designed to quantify glycemic variability by assessing fluctuation in glucose levels over time. It is particularly useful for identifying periods of significant glucose variability, which can be critical for managing diabetes effectively. The LI formula assesses variability in 3 consecutive glucose values at a time [4]. High LI indicates greater variability and potential instability in glucose levels, which can be associated with increased risks of hypoglycemia and hyperglycemia. Researchers compared the outcome of the HYPO score and the LI in patients living with type I diabetes who were undergoing islet transplant therapy [41]. The HYPO score is a measurement based on the frequency, severity, and degree of unawareness of a hypoglycemic event. Posttransplant patients experienced a significant reduction in the HYPO score and LI, indicating clinical relevance of LI as an objective measure for assessing the severity of glycemic lability [41]. The LI also had a high correlation with the metric, MAGE. This correlation makes it challenging to statistically isolate the unique

XSL•FO
RenderX

and independent contribution of liability to long-term risk assessment [58]. The LI also requires specialized software to address complex calculation methods.

Glucose Color Index

Standard CGM metrics are inadequate in representing the shape of glucose variability [42]. The glucose color index (GCI) is a new CGM metric that captures the shape of glucose variability. GCI uses periodogram signal decomposition. The metric was created through a multistep process involving the calculation of log-periodograms for each CGM time series, summarization using disjoint piece-wise linear models in 3 frequency regions, and refinement into a single value using canonical correlation analysis. The GCI is calculated by applying a linear combination of weights to the 6-parameter log-periodogram summaries, with these weights derived to maximize the test-retest correlation within participants over a 3-month period. Raw GCI values are then standardized to a scale of 0‑100 using a reference population for interpretability. In the study, the GCI was used to assess its reliability by comparing its 3-month test-retest correlation (R=.75) with standard CGM metrics and evaluate associations with diabetes comorbidities in older adults with type 2 diabetes [42]. The GCI showed significant associations (P<.05) with impaired physical functioning, frailty or prefrailty, cardiovascular disease, chronic kidney disease, and dementia or mild cognitive impairment after adjustment for confounders [42]. Overall, the study explains how the GCI captures clinically relevant information about glucose dysregulation that is not fully captured by standard CGM metrics. The GCI metric creates space for the development of CGM metrics that more fully incorporate time series information.

Overall, while metrics of glycemic variability provide granular insights into glucose stability and fluctuations that standard averages miss, their application is currently hindered by a lack of computational standardization and the frequent need for specialized software.

Composite Indices of Glycemic Quality and Risk

Glycemic Risk Index

The glycemic risk index (GRI) aims to assist with basic clinical interpretation of CGM data. To develop the GRI, researchers compiled a dataset of 14-day CGM tracings from 225 insulin-treated adults with diabetes [58]. Using a balanced incomplete block design, 330 clinicians with extensive experience in CGM analysis and interpretation ranked these tracings based on the quality of glycemia [58]. Principal component analysis and multiple regressions were used to create a model predicting clinician rankings using the following 7 standard metrics from an AGP: very low-glucose and low-glucose hypoglycemia, very high-glucose and high-glucose hyperglycemia, TIR, mean glucose, and CV. The resulting GRI provides a single-number summary of glycemic quality, with its hypoglycemia and hyperglycemia components offering actionable scores. Additionally, the GRI grid, a graphical display, enables clinicians and researchers to visualize and determine the glycemic effects of prescribed and investigational treatments, enhancing its utility in both clinical practice and research settings [43]. The GRI is based on clinician rankings rather than clinical outcomes. Additional studies are necessary to determine how accurately GRI predicts long-term complications. Furthermore, the GRI does not inherently distinguish glycemic risks by time of day (eg, daytime vs nighttime), which is crucial for making precise therapeutic adjustments [58].

M-Value

The M-value provides valuable insights into glycemic control and variability. It is defined as the difference between the observed blood sugar and normal blood sugar [37]. Lower M-values indicate better glycemic control, as demonstrated in a study comparing insulin glargine and insulin degludec in patients with type 1 diabetes using CGMs [44]. The researchers found that M-values were significantly smaller in patients treated with insulin degludec compared with those treated with insulin glargine, suggesting improved glycemic variability with insulin degludec. This study highlights the utility of the M-value in assessing the effectiveness of different insulin treatments and its role in evaluating glycemic control in clinical research [34]. A fundamental limitation is that the M-value is a hybrid measure that reflects both glucose variability and mean glycemia. It does not isolate variability itself [37]. Additionally, calculating the M-value requires complex calculation programs, which limits immediate application in acute settings like the ICU [46].

J-Index

The J-index was proposed as an improvement over the existing M-value, integrating both the mean blood glucose (MBG) and SD of blood glucose to provide a comprehensive representation of glycemic status [59]. By incorporating these 2 components, the J-index effectively captures both the mean level and variability of glycemia, addressing key aspects of glucose control. Moreover, 1 study applied the J-index to evaluate glycemic control in a group of patients with type 1 diabetes undergoing intensive insulin therapy [59]. Researchers used the J-index to assess the effectiveness of different insulin regimens and found it to be a valuable tool for comparing glycemic control between patient groups and monitoring individual progress over time [24,46,60]. Similar studies demonstrated the practical application of the J-index in clinical settings and its potential to provide meaningful insights into glucose management strategies for diabetic patients. Similar to the M-value, the J-index requires the use of specialized software or programming, thereby hindering its immediate application in clinical care settings.

GRADE

The GRADE score is an innovative metric designed to comprehensively evaluate glycemic control in patients with diabetes [22]. GRADE combines 2 risk quantities to summarize the overall risk associated with both hypoglycemia and hyperglycemia. The score's calculation involves a log-log transformation, with risk increasing as glucose levels deviate from an assigned nadir of 90 mg/dL. This approach provides a clinically meaningful measure of glycemic risk, applying different weights to hyperglycemic and hypoglycemic values.

The development of GRADE stemmed from the increasing practice of multiple daily glucose assessments in patients with

XSL•FO
RenderX

diabetes and the need for a more comprehensive evaluation of glycemic control [61]. To create this methodology, Hill et al [50] collaborated with 50 diabetes professionals who assigned risk values to a range of 40 blood glucose concentrations. From these responses, a generic function of glycemic risk was devised and applied to patient glucose profiles to generate an integrated risk score.

Interpreting GRADE scores is straightforward, with values below 5 corresponding to euglycemia. The simplicity of generating GRADE scores from any blood glucose profile makes it a valuable tool for clinicians. It can be used in conjunction with HbA_{1c} to provide a more comprehensive assessment of glycemic control, offering insights into the degree of risk associated with glycemic variability that HbA_{1c} alone cannot capture. The GRADE score does not directly measure glucose fluctuations. The derivation results in an expression that often reflects quasi-mean glycemia or a frequency distribution [37]. A high score can be generated by either hyperglycemia or hypoglycemia. Therefore, the formula does not explicitly specify the nature of the associated risk [37].

Average Daily Risk Range

The average daily risk range (ADRR) was created to serve as a single metric to summarize blood glucose variability [51]. The ADRR is a diabetes-specific measure designed to assess the risk of both hyperglycemia and hypoglycemia. It is calculated using mathematically transformed glucose levels to give equal weight to high and low blood glucose excursions [52]. ADRR requires at least 14 days of blood glucose data with a minimum of 3 checks per day, although some studies have used fewer days. The metric is scored based on risk categories: low risk (0 - 19), moderate risk (20-40), and high risk (40 and above) [62].

A study used ADRR to assess glycemic variability in children aged 2 - 6 years with type 1 diabetes [33]. The researchers found that 72% of the children's ADRR values fell into the "high-risk" range when using adult guidelines, suggesting a need for pediatric-specific ADRR standards [33]. The study also revealed that ADRR was highly correlated with indicators of hyperglycemia but only weakly correlated with measures of hypoglycemia in this young population [33]. Some articles in this review highlighted ADRR's broader applications in clinical settings [62]. It emphasized ADRR's ability to provide meaningful data on patients' risk for both hyperglycemia and hypoglycemia, offering insights not available from HbA_{1c} values alone. The review also noted that research has shown ADRR to be a reliable predictor of extreme blood glucose values across different diabetes types and patient ages.

Low Blood Glucose Index

Low blood glucose index (LBGI) was developed specifically to measure the risk associated with low blood glucose events. There are four distinct risk categories of the LBGI: (1) minimal risk for hypoglycemia (LBGI≤1.1), (2) low risk (1.1<LBGI≤2.5), (3) moderate risk (2.5<LBGI≤5), and (4) high risk (LBGI>5) [29]. The LBGI has been demonstrated to predict severe hypoglycemia, while the high blood glucose index (HBGI) is associated with HbA_{1c} levels and predicting hyperglycemic excursions. Additionally, the ADRR serves as a measure of overall glucose variability, capturing the risk of both hypoglycemia and hyperglycemia.

LBGI's primary strength is its ability to quantify the frequency and extent of low blood glucose excursions, placing more emphasis on the hypoglycemic risk [44]. It has been shown to be a robust predictor of symptomatic hypoglycemic episodes [54]. Because LBGI correlates well with the occurrence of severe hypoglycemia, it is a valuable tool for identifying people at higher risk of hypoglycemia episodes. However, LBGI only focuses on hypoglycemia and does not provide information about overall glycemic control or hyperglycemic events. Additionally, interpreting LBGI values may require a higher level of clinical experience as well as clinical context to translate into meaningful treatment decisions.

High Blood Glucose Index

High blood glucose index (HBGI) was designed to assess the risk associated with hyperglycemic events. Like LBGI, the HBGI quantifies glycemic variability focusing on hypoglycemic excursions. The cutoff points for the HBGI (4.5 and 9) are used to categorize the risk levels low, moderate, and high for hyperglycemia [56]. Risk scores above 0 are defined as HBGI [4]. In a study evaluating the improvement in glucose variability and its correlation with baseline HbA_{1c} and hypoglycemic events, the HBGI was found to have the largest proportional reduction (65.5%) among several glucose variability metrics after 24 weeks of treatment intensification. Furthermore, the HBGI was identified as the metric most predictive of improvements in HbA_{1c} levels [56].

This metric can be used to quantify the frequency and extent of high blood glucose excursions and can provide insights into the risk of hyperglycemia-related complications. HBGI has been shown to positively correlate with HbA_{1c}, making it useful for estimating overall glycemic control and potential long-term outcomes [57]. Similar to LBGI, a limitation of HBGI is that it provides a unidimensional view of glycemic control. Furthermore, the interpretation of HBGI values may not be intuitive for all people with diabetes and health care professionals, which may require additional education for effective use in clinical practice.

Index of Glycemic Control

Index of glycemic control (IGC) assesses the overall quality of glycemic control [23,24]. It is calculated as the sum of the hypoglycemia index and the hyperglycemia index, assigning more weight to severe hypo- or hyperglycemic values [22,24]. One study concluded the IGC to be one of the most sensitive criteria for detecting responses to therapeutic intervention, such as the unmasking of data [24]. A disadvantage of using IGC as a composite measure is that it relies solely on hypoglycemia and hyperglycemia components, which possibly excludes other valuable metrics [24].

Hypoglycemia Score

This metric is primarily based on a change in severe hypoglycemic events. Hypoglycemia score (HAS) aims to express overall glycemic control by simultaneously accounting

XSL•FO
RenderX

for 2 key factors: the incidence of hypoglycemia and the change in HbA_{1c} [22]. HAS provides a method for comprehensive assessment and does not require complex data or equations. Its current framework uses subjective weightings and lacks CIs, meaning the results are arbitrary and there is no assurance of reproducibility from one observer to the next [22].

Q-Score

The Q-score was developed to automate objective evaluations of CGM profiles and serve as a screening tool to identify patients who require therapeutic intervention [60]. It is calculated using MBG, TBR, TAR, within-day variability, and MODD for between-day variability. Q-score values range from poor (≥12.0) to very good (<4.0) [22]. The examination of components that contribute to the Q-score allows clinicians to identify weak points in glycemic control. The Q-score is similar to a difficulty rating for a task. The score significantly increases with the complexity of treatment. A limitation of the Q-score is that there is no evidence on its use of MODD offering a better insight into glycemic variability than standard measures, such as SD or CV [60].

Personal Glycemic State

Personal glycemic state (PGS)is designed to provide a simple measure for quickly identifying problem areas in glycemic control [22]. It is calculated using TIR, mean glucose, glycemic variability percentage, and the number of hypoglycemic episodes at <70 mg/dL and <54 mg/dL thresholds [22]. The score is used to complement HbA_{1c} by measuring glycemic control. However, a limitation is that it does not include hyperglycemia as a component, and considers the frequency of glucose oscillations, making it partially a measure of frequency stability rather than exclusively magnitude [22].

The Hypotriad

The hypotriad uses 3 components within a 3D space model to assess the severity and risk associated with hypoglycemia. Its components are area under the curve (AUC), time spent in hypoglycemia, and the frequency of hypoglycemic episodes per day [62]. These core metrics are integrated to calculate two related composite metrics known as (1) the intensity of hypoglycemia (IntHypo), defined as a vector of AUC and time, and (2) the hypoglycemia risk volume, which is the product of IntHypo and the episode rate [62]. A major limitation of the hypotriad is that it focuses exclusively on hypoglycemia and does not include any components describing hyperglycemia, which is an important factor for assessing overall glycemic control [22].

Glucose Pentagon and Comprehensive Glucose Pentagon

The glucose pentagon and comprehensive glucose pentagon (CGP) are composite metrics designed to provide a simultaneous numerical and visual representation of glycemic control [22]. The original glucose pentagon integrated 5 axes, including MBG, metrics for hyperglycemic AUC, time in hyperglycemia, SD, and HbA_{1c} [22]. The CGP modified this by eliminating HbA_{1c} and introducing mean glucose, %CV, time out of range, the intensity of hyperglycemia (IntHyper), and the intensity of hypoglycemia (IntHypo), with %CV replacing SD to account for abnormal glucose distributions and independence from mean glucose. The core output of the CGP is the area of the resulting pentagon, which increases as glycemic control worsens relative to a person without diabetes (area normalized to 1) [22]. A key limitation is the complexity of the nonstandard metrics like IntHyper and IntHypo, which use quadratic averages of AUC and duration, making them difficult for most clinicians to visualize and interpret [22].

CGM Index

Continuous glucose monitoring index (COGI) evaluates the quality of glucose control in patients with type 1 diabetes [22]. It is calculated using 3 weighted components: TIR, TBR, and SD, with TIR carrying the most weight [22]. COGI provides a focused view on the 3 components, also allowing for %CV to be used in replacement of SD to measure variability. A potential limitation of COGI is that its component weighting was determined by expert opinion and it does not account for TBR very low (levels below 54 mg/dL) [22].

Overall, composite indices of glycemic quality and risk attempt to simplify multidimensional CGM data into unified risk scores for easier comparison, yet their widespread clinical adoption is limited by complex, nonstandardized definitions and a frequent reliance on expert opinion rather than hard clinical outcomes.

Discordance Metrics: Acute Glycemic Gap

Acute glycemic gap ($GMI-HbA_{1c}$) is the difference between the GMI and HbA_{1c}. This metric was created to capture the discrepancy between a patient's recent glucose levels captured by CGMs and their HbA_{1c}. $GMI-HbA_{1c}$ is calculated by subtracting the actual HbA_{1c} from the GMI-derived CGM data. The study found that this metric was significantly associated with clinical outcomes in patients in the ICU. Specifically, it showed a positive correlation with the duration of ICU stay (r=0.533, P<.001) and was significantly higher in nonsurvivors at 28 days compared with survivors (mean 2.0, SD 1.3 vs mean 0.0, SD 1.6; P<.001) [46]. After adjusting for confounding factors, $GMI-HbA_{1c}$ remained an independent risk factor for longer ICU stay (coefficient=2.34, 95% CI 0.54 - 4.14; P=.02) and higher 28-day mortality rate (hazard ratio 2.42, 95% CI 1.01 - 5.76; P=.046). The authors concluded that $GMI-HbA_{1c}$ is an independent risk factor for longer ICU stay and 28-day mortality rate [46], suggesting that CGM might be beneficial for critically ill patients in the ICU setting, regardless of whether they have diabetes.

By quantifying the discrepancy between CGM-derived estimates and laboratory biomarkers, discordance metrics can help identify independent risk factors for adverse outcomes and highlight critical nonglycemic physiological variances that impact assessment.

Discussion

This narrative review mapped the current landscape of 36 CGM metrics, providing a comprehensive resource to navigate between CGM metric appropriateness for clinical simplicity and research complexity. The standard clinical metrics, such as TIR and GMI, have achieved widespread adoption due to their

XSL•FO
RenderX

simplicity, interpretability, and direct correlation with established clinical metrics such as HbA_{1c} as well as long-term complications. These metrics can be used for daily glucose management decisions and are ideal for use in time-constrained clinical settings, offering easy-to-digest snapshots of glycemic control. However, these metrics may mask underlying glycemic variability and fail to identify more granular patterns that could improve patient outcomes.

The expanding ecosystem of emerging and composite metrics identifies and summarizes glycemic patterns using composite risk scoring (eg, GRI, GRADE, and ADRR) and granular analysis (eg, MAGE, MODD, and CONGA). These metrics involve more complex formulas that offer mechanistic insights, which allow researchers to characterize glycemic health more fully, offering insight into daily fluctuations. This increased granularity often comes with increased complexity. Developing intuitive visualization and standard reports of such metrics, such as the existing AGP, has the potential to increase the usage of these metrics in the future.

A critical finding of this review is the lack of standardization in how complex metrics are calculated. Even well-known metrics, such as MAGE, are highly sensitive to data processing decisions, such as the specific software package used. Inherent differences in CGM sensor hardware, including differences in sampling frequency (eg, 1-min vs 5-min intervals), add to the complexity of standardizing metric calculation. This variability means that scores in one study may not be comparable with another. This lack of standardization and reproducibility is a major barrier to universal metric adoption and undermines the validity of CGM research. To address this, the field must move toward open science frameworks [21] and adoption of standardized code libraries [25-28] to ensure that definitions and interpretations of metrics are consistent across devices and trials. Consensus standardization, implementation, and interpretation of such metrics will need to be established before their potential for impact can be fully realized.

This review has several limitations. First, the search strategy was restricted to articles published in English and indexed in PubMed or Google Scholar. Studies published in other languages or limited to other databases may have been missed. Second, our primary search terms included the 27 metrics identified by the DRH [21]. This prespecified list may exclude niche or newly proposed metrics from this quickly evolving landscape that have not yet achieved broader recognition. Third, the studies included were published across nearly a 25-year period. CGM hardware and software have evolved dramatically in this time. Articles included therefore used a heterogeneous mix of CGM devices ranging from early-generation sensors to contemporary systems. Evolving sensor accuracy and sampling frequency across these generations can influence the calculation and precision of the reported metrics, making assessment of metrics challenging. Furthermore, the increase of complex metrics can unknowingly widen health disparities, particularly for patients and nonspecialist clinicians. While this narrative review highlights the breadth of available CGM metrics, many of these metrics require specialized knowledge, software, and/or large amounts of data for interpretation. This requirement poses a significant barrier for individuals with low digital or health literacy. Finally, this review focused primarily on the definitions and calculations of metrics. An evaluation of the strength of evidence linking each emerging metric to hard clinical outcomes was beyond the scope of this paper.

As CGM technology continues to advance and become more accessible, the role of CGM metrics in diabetes management is likely to grow. Future research may focus on refining existing metrics, developing new ones, and further elucidating their relationships with clinical outcomes. While consensus exists for core clinical metrics, the lack of standardization for more complex metrics remains a barrier to universal translation. The value of CGM metrics lies in their ability to provide reliable day-to-day insights, playing a crucial role in breakthrough discoveries and personalized glucose management for individuals living with diabetes. Aiming toward standardization of definitions, calculations, and methods of usage for CGM metrics contributes to the development of an open science framework ensuring these tools are accessible and actionable for participants who value these metrics.

Acknowledgments

We acknowledge the valuable feedback and scholarly guidance provided by Narayanan Krishnamurthi, PhD, during the preparation of this literature review. We also acknowledge the support of the Diabetes Research Hub Team, whose resources facilitated the research process.

Generative AI was used to assist in the article screening phase to assess the relevance and adequacy of materials for this review. The authors have reviewed and edited AI-generated material to ensure accuracy and take full responsibility for the integrity and final version of the manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

This study is a literature review and does not involve the generation or analysis of primary data. Therefore, no dataset is available.

XSL•FO
RenderX

Authors' Contributions

EW Jr: conceptualization, data curation, writing – original draft, writing – review & editing
CR: writing – original draft, writing – review & editing
AA: data curation, writing – review & editing
CH: data curation, writing – review & editing
AD: data curation, writing – review & editing
DK: conceptualization, writing – review & editing, funding acquisition
JE: conceptualization, writing – original draft, writing – review & editing, funding acquisition

Conflicts of Interest

DK is a consultant for Afon, embecta, Glooko, Glucotrack, Lifecare, Sanofi, SynchNeuro, and Thirdwayv. JE receives federal funding from the Food and Drug Administration, National Institute on Minority Health and Health Disparities, and National Center for Advancing Translational Sciences and is a consultant for Sanofi, Glooko, and Dexcom.

Checklist 1
PRISMA Checklist
[DOCX File, 276 KB - diabetes_v11i1e92455_app1.docx]

References

1. Ong KL, Stafford LK, McLaughlin SA, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet 2023 Jul;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6]
2. Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. Biomark Insights 2016;11:95-104. [doi: 10.4137/BMI.S38440] [Medline: 27398023]
3. Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. Diabetes Technol Ther 2016 Feb;18 Suppl 2(Suppl 2):S3-S13. [doi: 10.1089/dia.2015.0417] [Medline: 26784127]
4. Hill NR, Oliver NS, Choudhary P, Levy JC, Hindmarsh P, Matthews DR. Normal reference range for mean tissue glucose and glycemic variability derived from continuous glucose monitoring for subjects without diabetes in different ethnic groups. Diabetes Technol Ther 2011 Sep;13(9):921-928. [doi: 10.1089/dia.2010.0247] [Medline: 21714681]
5. Maiorino MI, Di Martino N, Angelino S, et al. The therapeutic efficacy of continuous glucose monitoring in diabetes: an updated meta-analysis with meta-regression. Endocrine 2025 Dec;90(3):1138-1148. [doi: 10.1007/s12020-025-04405-6] [Medline: 40952636]
6. Klonoff DC, Ahn D, Drincic A. Continuous glucose monitoring: a review of the technology and clinical use. Diabetes Res Clin Pract 2017 Nov;133:178-192. [doi: 10.1016/j.diabres.2017.08.005] [Medline: 28965029]
7. Martens T, Beck RW, Bailey R, et al. Effect of continuous glucose monitoring on glycemic control in patients with type 2 diabetes treated with basal insulin: a randomized clinical trial. JAMA 2021 Jun 8;325(22):2262-2272. [doi: 10.1001/jama.2021.7444] [Medline: 34077499]
8. Alfadli SF, Alotaibi YS, Aqdi MJ, et al. Effectiveness of continuous glucose monitoring systems on glycemic control in adults with type 1 diabetes: a systematic review and meta-analysis. Metabol Open 2025 Sep;27:100382. [doi: 10.1016/j.metop.2025.100382] [Medline: 40791933]
9. Steck AK, Dong F, Geno Rasmussen C, et al. CGM metrics predict imminent progression to type 1 diabetes: autoimmunity screening for kids (ASK) study. Diabetes Care 2022 Feb 1;45(2):365-371. [doi: 10.2337/dc21-0602] [Medline: 34880069]
10. Phillip M, Achenbach P, Addala A, et al. Consensus guidance for monitoring individuals with islet autoantibody-positive pre-stage 3 type 1 diabetes. Diabetes Care 2024 Aug 1;47(8):1276-1298. [doi: 10.2337/dci24-0042] [Medline: 38912694]
11. American Diabetes Association Professional Practice Committee. 7. Diabetes technology: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S146-S166. [doi: 10.2337/dc25-S007] [Medline: 39651978]
12. Rodbard D. Glucose variability: a review of clinical applications and research developments. Diabetes Technol Ther 2018 Jun;20(S2):S25-S215. [doi: 10.1089/dia.2018.0092] [Medline: 29916742]
13. Schnell O, Barnard K, Bergenstal R, et al. Role of continuous glucose monitoring in clinical trials: recommendations on reporting. Diabetes Technol Ther 2017 Jul;19(7):391-399. [doi: 10.1089/dia.2017.0054] [Medline: 28530490]
14. Ettestad TM, Willis HJ, Johnson E, Green CR, Bergenstal RM. Continuous glucose monitoring (CGM) as a behavior modification tool: a case series on CGM-guided nutrition in people with type 2 diabetes not using insulin. Clin Diabetes 2025;43(4):610-617. [doi: 10.2337/cd24-0106] [Medline: 41140737]
15. Richardson KM, Jospe MR, Bohlen LC, Crawshaw J, Saleh AA, Schembre SM. The efficacy of using continuous glucose monitoring as a behaviour change tool in populations with and without diabetes: a systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2024 Dec 23;21(1):145. [doi: 10.1186/s12966-024-01692-6] [Medline: 39716288]

XSL•FO
RenderX

16. Dunn TC, Ajjan RA, Bergenstal RM, Xu Y. Is it time to move beyond TIR to TITR? Real-world data from over 20,000 users of continuous glucose monitoring in patients with type 1 and type 2 diabetes. Diabetes Technol Ther 2024 Mar;26(3):203-210. [doi: 10.1089/dia.2023.0565]
17. Spanakis EK, Cook CB, Kulasa K, et al. A consensus statement for continuous glucose monitoring metrics for inpatient clinical trials. J Diabetes Sci Technol 2023 Nov;17(6):1527-1552. [doi: 10.1177/19322968231191104] [Medline: 37592726]
18. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
19. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023 Jan;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9]
20. Pleus S, Eichenlaub M, Eriksson Boija E, et al. The need for standardization of continuous glucose monitoring performance evaluation: an opinion by the international federation of clinical chemistry and laboratory medicine working group on continuous glucose monitoring. J Diabetes Sci Technol 2026 Jan;20(1):201-206. [doi: 10.1177/19322968241296097] [Medline: 39540398]
21. Ayers AT, Ho CN, Friedman E, Espinoza J, Shah SN, Klonoff DC. How the diabetes research hub will modernize and enhance diabetes data utilization. J Diabetes Sci Technol 2026 Jan;20(1):3-6. [doi: 10.1177/19322968241306129] [Medline: 39707843]
22. Nguyen M, Han J, Spanakis EK, Kovatchev BP, Klonoff DC. A review of continuous glucose monitoring-based composite metrics for glycemic control. Diabetes Technol Ther 2020 Aug;22(8):613-622. [doi: 10.1089/dia.2019.0434] [Medline: 32069094]
23. Rodbard D. The challenges of measuring glycemic variability. J Diabetes Sci Technol 2012 May 1;6(3):712-715. [doi: 10.1177/193229681200600328] [Medline: 22768904]
24. Rodbard D, Bailey T, Jovanovic L, Zisser H, Kaplan R, Garg SK. Improved quality of glycemic control and reduced glycemic variability with use of continuous glucose monitoring. Diabetes Technol Ther 2009 Nov;11(11):717-723. [doi: 10.1089/dia.2009.0077] [Medline: 19905888]
25. Broll S, Urbanek J, Buchanan D, et al. Interpreting blood glucose data with R package iglu. PLOS ONE 2021;16(4):e0248560. [doi: 10.1371/journal.pone.0248560] [Medline: 33793578]
26. Vigers T, Chan CL, Snell-Bergeon J, et al. cgmanalysis: an R package for descriptive analysis of continuous glucose monitor data. PLoS ONE 2019;14(10):e0216851. [doi: 10.1371/journal.pone.0216851] [Medline: 31603912]
27. Zhang XD, Zhang Z, Wang D. CGManalyzer: an R package for analyzing continuous glucose monitoring studies. Bioinformatics 2018 May 1;34(9):1609-1611. [doi: 10.1093/bioinformatics/btx826] [Medline: 29315360]
28. Millard LAC, Patel N, Tilling K, Lewcock M, Flach PA, Lawlor DA. GLU: a software package for analysing continuously measured glucose levels in epidemiology. Int J Epidemiol 2020 Jun 1;49(3):744-757. [doi: 10.1093/ije/dyaa004] [Medline: 32737505]
29. Clarke W, Kovatchev B. Statistical tools to analyze continuous glucose monitor data. Diabetes Technol Ther 2009 Jun;11 Suppl 1(Suppl 1):S45-S54. [doi: 10.1089/dia.2008.0138] [Medline: 19469677]
30. Wybrańczyk Ł, Brudzińska A, Jarosz-Chobot P, Deja G. Glucose management indicator – potential factors affecting differences in comparison with HbA1c and clinical significance of this phenomenon. pedm 2023;29(3):190-195. [doi: 10.5114/pedm.2023.130028]
31. Bergenstal RM, Beck RW, Close KL, et al. Glucose Management Indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. Diabetes Care 2018 Nov;41(11):2275-2280. [doi: 10.2337/dc18-1581] [Medline: 30224348]
32. Shah VN, Xu Y, Dabiri Y, Mohanadas HP, Cheng A, Dunn TC. Association of HbA1c and an updated glucose management indicator (uGMI) with incident diabetic retinopathy in adults with type 1 diabetes: a longitudinal study. Diabetologia 2026 Mar;69(3):610-617. [doi: 10.1007/s00125-025-06599-w] [Medline: 41217525]
33. Monaghan M, Younge TB, McCarter R, Cogen FR, Streisand R. Average Daily Risk Range (ADRR) in young children with type 1 diabetes. J Diabetes Sci Technol 2014 Jan;8(1):70-73. [doi: 10.1177/1932296813511734] [Medline: 24876540]
34. Hamasaki H. Comparison of glycemic variability by using insulin glargine and insulin degludec in Japanese patients with type 1 diabetes, monitored by continuous glucose monitoring: a preliminary report. J Endocrinol Metab 2013. [doi: 10.4021/jem193w]
35. Rawla P, Vellipuram AR, Bandaru SS, Pradeep Raj J. Euglycemic diabetic ketoacidosis: a diagnostic and therapeutic dilemma. Endocrinol Diabetes Metab Case Rep 2017;2017:17-0081. [doi: 10.1530/EDM-17-0081] [Medline: 28924481]
36. Vergès B, Pignol E, Rouland A, et al. Glycemic variability assessment with a 14-day continuous glucose monitoring system: when and how long to measure MAGE (Mean Amplitude of Glucose Excursion) for optimal reliability? J Diabetes Sci Technol 2022 Jul;16(4):982-987. [doi: 10.1177/1932296821992060] [Medline: 33567877]
37. Service FJ. Glucose variability. Diabetes 2013 May;62(5):1398-1404. [doi: 10.2337/db12-1396] [Medline: 23613565]
38. McDonnell CM, Donath SM, Vidmar SI, Werther GA, Cameron FJ. A novel approach to continuous glucose analysis utilizing glycemic variation. Diabetes Technol Ther 2005 Apr;7(2):253-263. [doi: 10.1089/dia.2005.7.253] [Medline: 15857227]

XSL•FO
RenderX

39. Reddy M, Godsland IF, Barnard KD, et al. Glycemic variability and Its impact on quality of life in adults with type 1 diabetes. J Diabetes Sci Technol 2015 Aug 18;10(1):60-66. [doi: 10.1177/1932296815601440] [Medline: 26285951]
40. Kusunoki Y, Konishi K, Tsunoda T, Koyama H. Significance of glycemic variability in diabetes mellitus. Intern Med 2022 Feb 1;61(3):281-290. [doi: 10.2169/internalmedicine.8424-21] [Medline: 34544956]
41. Ryan EA, Shandro T, Green K, et al. Assessment of the severity of hypoglycemia and glycemic lability in type 1 diabetic subjects undergoing islet transplantation. Diabetes 2004 Apr;53(4):955-962. [doi: 10.2337/diabetes.53.4.955] [Medline: 15047610]
42. Sartini J, Fang M, Rooney MR, Selvin E, Coresh J, Zeger S. Glucose color index: development and validation of a novel measure of the shape of glycemic variability. J Diabetes Sci Technol 2025 Nov;19(6):1505-1514. [doi: 10.1177/19322968241245654] [Medline: 38641966]
43. Kim JY, Yoo JH, Kim JH. Comparison of glycemia risk index with time in range for assessing glycemic quality. Diabetes Technol Ther 2023 Dec;25(12):883-892. [doi: 10.1089/dia.2023.0264]
44. Kovatchev B, Meng Z, Cali AMG, Perfetti R, Breton MD. Low blood glucose index and hypoglycaemia risk: insulin glargine 300 U/mL versus insulin glargine 100 U/mL in type 2 diabetes. Diabetes Ther 2020 Jun;11(6):1293-1302. [doi: 10.1007/s13300-020-00808-y] [Medline: 32304086]
45. Brett McQueen R, Perez-Nieves M, Todd Alonso G, et al. Association between continuous glucose monitoring metrics and clinical outcomes in adults with type 1 diabetes in a real-world setting. Diabetes Res Clin Pract 2024 Jun;212:111690. [doi: 10.1016/j.diabres.2024.111690] [Medline: 38697300]
46. Ha EY, Chung SM, Park IR, Lee YY, Choi EY, Moon JS. Novel glycemic index based on continuous glucose monitoring to predict poor clinical outcomes in critically ill patients: a pilot study. Front Endocrinol (Lausanne) 2022;13:869451. [doi: 10.3389/fendo.2022.869451] [Medline: 35600594]
47. Wójcicki JM. “J”-index. A new proposition of the assessment of current glucose control in diabetic patients. Horm Metab Res 1995 Jan;27(1):41-42. [doi: 10.1055/s-2007-979906] [Medline: 7729793]
48. Zhou J, Chen Z, Huang HN, Ou CQ, Li X. Association between various blood glucose variability-related indicators during early ICU admission and 28-day mortality in non-diabetic patients with sepsis. Diabetol Metab Syndr 2025 Jan 20;17(1):22. [doi: 10.1186/s13098-025-01580-4] [Medline: 39828689]
49. Wojcicki JM, Ladyzynski P, Krzymien J, et al. What we can really expect from telemedicine in intensive diabetes treatment: results from 3-year study on type 1 pregnant diabetic women. Diabetes Technol Ther 2001;3(4):581-589. [doi: 10.1089/15209150152811207] [Medline: 11911170]
50. Hill NR, Hindmarsh PC, Stevens RJ, Stratton IM, Levy JC, Matthews DR. A method for assessing quality of control from glucose profiles. Diabet Med 2007 Jul;24(7):753-758. [doi: 10.1111/j.1464-5491.2007.02119.x] [Medline: 17459094]
51. Kovatchev BP, Otto E, Cox D, Gonder-Frederick L, Clarke W. Evaluation of a new measure of blood glucose variability in diabetes. Diabetes Care 2006 Nov;29(11):2433-2438. [doi: 10.2337/dc06-1085] [Medline: 17065680]
52. Kovatchev B, Umpierrez G, DiGenio A, Zhou R, Inzucchi SE. Sensitivity of traditional and Risk-Based Glycemic Variability Measures to the effect of glucose-lowering treatment in type 2 diabetes mellitus. J Diabetes Sci Technol 2015 Jun 15;9(6):1227-1235. [doi: 10.1177/1932296815587014] [Medline: 26078255]
53. Patton SR, Clements MA. Average daily risk range as a measure for clinical research and routine care. J Diabetes Sci Technol 2013 Sep 1;7(5):1370-1375. [doi: 10.1177/193229681300700529] [Medline: 24124966]
54. Crenier L, Abou-Elias C, Corvilain B. Glucose variability assessed by low blood glucose index is predictive of hypoglycemic events in patients with type 1 diabetes switched to pump therapy. Diabetes Care 2013 Aug;36(8):2148-2153. [doi: 10.2337/dc12-2058] [Medline: 23404296]
55. Fabris C, Patek SD, Breton MD. Are risk indices derived from CGM interchangeable with SMBG-based indices? J Diabetes Sci Technol 2016 Jan;10(1):50-59. [doi: 10.1177/1932296815599177]
56. Kovatchev BP. Metrics for glycaemic control - from HbA1c to continuous glucose monitoring. Nat Rev Endocrinol 2017 Jul;13(7):425-436. [doi: 10.1038/nrendo.2017.3] [Medline: 28304392]
57. Sartore G, Chilelli NC, Burlina S, et al. The importance of HbA1c and glucose variability in patients with type 1 and type 2 diabetes: outcome of continuous glucose monitoring (CGM). Acta Diabetol 2012 Dec;49 Suppl 1(S1):S153-S160. [doi: 10.1007/s00592-012-0391-4] [Medline: 22466072]
58. Klonoff DC, Wang J, Rodbard D, et al. A Glycemia Risk Index (GRI) of hypoglycemia and hyperglycemia for continuous glucose monitoring validated by clinician ratings. J Diabetes Sci Technol 2023 Sep;17(5):1226-1242. [doi: 10.1177/19322968221085273]
59. Vigersky RA. Escaping the hemoglobin A1c-centric world in evaluating diabetes mellitus interventions. J Diabetes Sci Technol 2015 Feb 19;9(5):1148-1151. [doi: 10.1177/1932296815576363] [Medline: 25697718]
60. Augstein P, Heinke P, Vogt L, et al. Q-Score: development of a new metric for continuous glucose monitoring that enables stratification of antihyperglycaemic therapies. BMC Endocr Disord 2015 May 1;15(1):22. [doi: 10.1186/s12902-015-0019-0] [Medline: 25929322]
61. Thomas A, Shin J, Jiang B, McMahon C, Kolassa R, Vigersky RA. The development of new composite metrics for the comprehensive analytic and visual assessment of hypoglycemia using the hypo-triad. J Diabetes Sci Technol 2018 Jan;12(1):69-75. [doi: 10.1177/1932296817721242] [Medline: 28748706]

XSL•FO
RenderX

62. Vigersky RA, Shin J, Jiang B, Siegmund T, McMahon C, Thomas A. The comprehensive glucose pentagon: a glucose-centric composite metric for assessing glycemic control in persons with diabetes. J Diabetes Sci Technol 2018 Jan;12(1):114-123. [doi: 10.1177/1932296817718561] [Medline: 28748705]
63. Beck RW, Connor CG, Mullen DM, Wesley DM, Bergenstal RM. The fallacy of average: how using HbA1c alone to assess glycemic control can be misleading. Diabetes Care 2017 Aug;40(8):994-999. [doi: 10.2337/dc17-0636] [Medline: 28733374]

Abbreviations

%CV: glucose variability
ADRR: average daily risk range
AGP: ambulatory glucose profile
AUC: area under the curve
CGM: continuous glucose monitoring
CGP: comprehensive glucose profile
COGI: continuous glucose monitoring index
CONGA: continuous overall net glycemic action
CV: coefficient of variation
DRH: Diabetes Research Hub
GCI: glucose color index
GMI: glucose management indicator
GMI−HbA$_{1c}$: acute glycemic gap
GRADE: glycemic risk assessment diabetes equation
GRI: glycemic risk index
HAS: hypoglycemia score
HbA$_{1c}$: hemoglobin A$_{1c}$
HBGI: high blood glucose index
IGC: index of glycemic control
IntHyper: intensity of hyperglycemia
IntHypo: intensity of hypoglycemia
LBGI: low blood glucose index
LI: lability index
MAGE: mean average glucose excursions
MBG: mean blood glucose
MODD: mean of daily differences
TAR: time above range
TBR: time below range
TIR: time in range
uGMI: updated glucose management indicator

Edited by I Steenstra; submitted 30.Jan.2026; peer-reviewed by M Grady, S Harper; revised version received 22.May.2026; accepted 25.May.2026; published 23.Jul.2026.

Please cite as:
Williams Jr E, Rand C, Ayers A, Ho C, Dunova A, Klonoff D, Espinoza J
The Evolving Landscape of Continuous Glucose Monitoring Metrics in Type 1 Diabetes: Narrative Literature Review
JMIR Diabetes 2026;11:e92455
URL: https://diabetes.jmir.org/2026/1/e92455
doi:10.2196/92455

© Eric Williams Jr, Casey Rand, Alessandra Ayers, Cindy Ho, Ashley Dunova, David Klonoff, Juan Espinoza. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Technologies for Diabetic Foot Ulcer Monitoring and Risk Prediction: Systematic Review

Houriyeh Ehtemam[1], MSc; Simin Salehinejad[2], PhD; Dan Robbins[3], PhD; Alireza Sanaei[1], PhD; Hassan Shirvani[1], PhD; Shabnam Sadeghi-Esfahlani[1], PhD

[1]School of Engineering and the Built Environment, Faculty of Science and Engineering, Anglia Ruskin University, Essex, Chelmsford, United Kingdom

[2]Endocrinology and Metabolism Research Center, Kerman University of Medical Sciences, Endocrinology and Methabolism Research Center, Jahad Bolvd, Ebnsina Ave, Kerman, Iran

[3]Medical Technology Research Centre, School of Allied Health, Faculty of Health, Education, Medicine & Social Care, Anglia Ruskin University, Essex, Chelmsford, United Kingdom

Corresponding Author:
Simin Salehinejad, PhD
Endocrinology and Metabolism Research Center, Kerman University of Medical Sciences, Endocrinology and Methabolism Research Center, Jahad Bolvd, Ebnsina Ave, Kerman, Iran

Abstract

Background: Wearable technologies, including smart insoles and sensor-equipped footwear, enable continuous monitoring of key foot parameters such as plantar pressure and temperature in individuals at risk of diabetic foot ulcers (DFUs).

Objective: This systematic review aimed to evaluate the technological characteristics and clinical applications of wearable devices for monitoring DFU-related parameters.

Methods: This review was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Journal articles, theses, and dissertations evaluating wearable technologies for DFU prevention or monitoring were eligible if they involved human participants and were published in English or Persian. Studies focused on nonwearable or non-foot–based systems were excluded. A comprehensive search was conducted in PubMed, Embase, Web of Science, and Scopus from July 2024 to May 2025. Two reviewers (HE and SS) independently screened studies and extracted data. The methodological quality of included studies was assessed using the Mixed Methods Appraisal Tool (MMAT) 2018. Results were synthesized using descriptive synthesis.

Results: A total of 1088 records were identified, of which 23 studies met the inclusion criteria. The included studies varied in design, sample size, and follow-up duration. Wearable devices included smart insoles, socks, and external sensors, primarily monitoring plantar pressure and temperature. Devices differed in sensor type, placement, number, communication protocols, and data acquisition rates. Participants typically had diabetes, and many had a history of neuropathy or prior DFUs.

Conclusions: Wearable technologies show promise for monitoring DFU risk factors and supporting early detection. However, the evidence base is limited by heterogeneity in study designs, small sample sizes, and short follow-up periods. Further high-quality studies are required to evaluate their potential clinical benefits, long-term outcomes, and role in preventing DFUs and improving patient care.

(JMIR Diabetes 2026;11:e84463) doi:10.2196/84463

KEYWORDS

diabetic foot ulcer; DFU prevention; foot monitoring; wearable technology; smart insole; Internet of things

Introduction

Globally, diabetes (both type 1 and type 2) affects approximately 463 million individuals. According to the International Diabetes Federation, this figure is projected to rise to 578 million by 2030 and reach 700 million by 2045 [1]. Diabetes is associated with different medical comorbidities over an extended duration. Diabetic foot ulcers (DFUs) are a common and debilitating complication of diabetes mellitus that can lead to significant morbidity and mortality [2]. DFU is a complex, multifactorial condition that often develops from a combination of peripheral vascular disease, neuropathy, immune dysfunction, Charcot's neuroarthropathy, previous foot ulceration, and osteomyelitis [3,4]. Consequently, DFUs can lead to nerve damage, resulting in reduced pain sensation and leaving patients unaware of injuries or ulcers. The development of complications can start up to 5 years before a diagnosis of diabetes [5,6].

Approximately 5%-7% of individuals with diabetes are estimated to have experienced, or be experiencing foot ulceration at the time of this writing. In Scotland, a reported 2.5% of the

XSL•FO
RenderX

diagnosed diabetic population had active foot ulcers in December 2010 [7,8]. Leading the National Health Service (NHS) to allocate around £1 billion (GBP £1=US $1.29 as of December 31, 2017; approximately US $1.29 billion) annually toward addressing foot care needs of individuals with diabetes [9].

These trends highlight the urgent need for improved DFU prevention strategies [10].

Preventing diabetic foot complications involves a multifaceted approach. This includes lifestyle management, optimizing metabolic control by regulating glucose levels, identifying and screening individuals at high risk for DFUs, and providing patient education to promote foot self-examination and foot care knowledge [11]. Patients are educated to routinely inspect their feet for hyperkeratosis (thickening of the skin's outer layer), fungal infections, skin lesions, and foot deformities [12]. Footwear control is also emphasized, as repeated trauma from tight or poorly fitting shoes can act as a triggering factor [13]. Achieving this goal can be facilitated by developing affordable and effective wearable devices [14,15]. Wearable technology refers to devices designed to be worn as accessories that facilitate user interaction through both physical inputs and nonphysical modalities, such as voice-based commands [16]. A specific subset of these, referred to as in-shoe wearables, includes devices embedded within or inserted into footwear (eg, smart insoles) for monitoring foot-related parameters such as plantar pressure and temperature. In contrast, nearable devices refer to systems that are not worn on the body but are used in close proximity, such as thermal mats or foot scanners used in clinical or home settings [17]. Recent advances in wearable health care technologies have also been highlighted in broader biomedical applications. A recent review of smart wearable orthoses and sensor-integrated biomedical devices emphasized the growing role of intelligent sensing systems in continuous health monitoring, rehabilitation, and personalized care [18]. The application of skin temperature monitoring represents a promising approach to reduce the risk of DFUs [19]. For example, a clinical trial study conducted by Ming et al [20] has highlighted the significance of thermal imaging techniques in detecting early abnormalities in foot temperature patterns associated with DFUs [20]. Plantar pressure devices are used for monitoring diabetic foot-related issues by measuring weight distribution while standing or walking, which is beneficial for monitoring foot health, especially for people with diabetes [21]. Several studies have addressed the usability of these devices; that is, Gupta et al [22] proposed an insole with high-resolution sensing, equipped with 126 sensing nodes [23]. Their study assessed the functionality of this insole in both static and dynamic conditions, with results confirming its usability for predicting diabetic ulceration. Additionally, the assessment of shear stress (a force that acts parallel to a surface) emerges as another valuable metric for monitoring foot conditions, particularly in predicting the occurrence of DFUs [24].

This study aims to systematically review wearable technologies used in DFU monitoring, focusing on their technical characteristics and reported clinical outcomes.

Methods

Overview

Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we systematically identified, selected, and evaluated relevant studies for our review [25]. No protocol was registered for this review.

Ethical Considerations

Given the nature of the review analyzing secondary data, no ethics approval was required.

Research Strategy

We searched the following databases from July 2024 to May 2025: PubMed, Embase, Web of Science, and Scopus for the English language documents, as well as databases including Scientific Information Database (SID) in Persian.

Search terms were combined using appropriate Boolean operators and included subject heading terms/keywords for 2 key themes (Medical Subject Headings [MeSH] terms), adjusted for each database. The 2 main terms were DFUs (eg, "diabetic foot" OR "diabetic foot ulcer," OR "DFU," OR "foot modeling") AND sensors (eg, "wearable device," OR "wearable computers," OR "electronic skin," OR "wearable technology," OR "wearable electronic device," OR "fabric sensor," OR "smart sock," OR "smart insole," OR "remote sensing," OR "flexible sensor"). Following the database searches, additional manual searching was performed. The records obtained through the database search were managed using EndNote X7 (Clarivate).

To define the research question, the PICO (Population, Intervention, Comparison, and Outcome) framework was applied as follows:

- Population (P): individuals with diabetes or at risk of DFUs;
- Intervention (I): wearable technologies for monitoring foot-related parameters (eg, plantar pressure and temperature);
- Comparison (C): variations in study characteristics and device features;
- Outcome (O): reported technical characteristics and clinical outcomes.

Based on this framework, the research question guiding this review was: "What are the key technological characteristics and clinical outcomes of wearable devices used for monitoring diabetic foot ulcers?"

Inclusion Criteria

The inclusion criteria focused on studies published as journal articles, theses, and dissertations. The following criteria guided study selection: (1) the study evaluated wearable devices, either commercially available or researcher-developed, used for foot monitoring, including in-shoe systems such as smart insoles, sensor-integrated footwear, smart socks, or similar wearable sensor devices that measured parameters such as plantar pressure, temperature, or humidity; (2) published in English or Persian; (3) it involved human participants rather than laboratory animals; and (4) it addressed either the prevention of DFUs (eg, monitoring high-risk individuals with neuropathy or a history

XSL•FO
RenderX

of prior ulcers) or the monitoring of active DFUs (eg, wound progression and healing). Studies were excluded if they did not use systems or sensors attached to the inside of the shoe or inserted, such as insoles or other removable devices, to monitor DFUs.

Study Selection

After removing duplicates, the first and second co-authors (HE and SS) independently screened titles and abstracts to remove irrelevant studies. To confirm eligibility, studies passing the title-abstract screening were examined by the first and second authors (HE and SS) at the full-text screening phase. In the last phase, the full texts of articles were assessed again by the first and second authors (HE and SS) to determine their originality and to extract data. Interrater reliability was assessed during the full-text screening stage using Cohen kappa [26]. All screening phases were conducted independently using EndNote and Excel (Microsoft Corp).

Data Extraction

Data extraction was completed independently by the first and second authors (HE and SS). When needed, the first and second authors (HE and SS) checked their data extraction process with the supervisor for accuracy and clarifications. Data extraction was conducted using Microsoft Excel 2016 spreadsheet. The following information was extracted from each included study: (1) study characteristics, including authors of the study, year of data publication, study location, study design, and sample size; (2) participant characteristics, including type of diabetes (if mentioned), follow-up period, and history of DFUs and neuropathy; (3) technical characteristics, including application of wearable devices, key metrics, sensors' location, count of sensors, sensors' technology, communication method, range of measurement, acquisition rate or sampling frequency, and user interface; and (4) clinical outcomes. When information was not directly available, the corresponding authors (SS) were contacted to obtain additional details.

Quality Assessment

The quality of the included articles was assessed using the Mixed Methods Appraisal Tool (MMAT 2018) following the search process, which includes a checklist to appraise the methodological quality for qualitative, quantitative, and specifically mixed methods studies. MMAT's 5 sets of criteria were adopted to evaluate the quality of each type of study included in our analysis, namely qualitative, randomized controlled trial (RCT), nonrandomized, quantitative, descriptive, and mixed methods studies [27]. This rigorous assessment process allowed us to evaluate the included studies' methodological quality and ensure our findings' reliability and validity (Table S1 in Multimedia Appendix 1 [3,19,20,22,28-68]).

Bias Management

To limit selection bias, we did not apply any restrictions regarding the type of study or the population studied, allowing for the inclusion of a broad and diverse range of studies. To reduce publication bias, we excluded previously published review and meta-analysis papers, focusing instead on primary studies to ensure the originality of the included data. In addition to peer-reviewed journal articles, gray literature sources such as theses and dissertations were not excluded (if available) to minimize publication bias and capture a broader range of evidence. These sources should be identified through database searches and screened using the same eligibility criteria as published studies.

The review included studies published in both English and Persian to enhance the comprehensiveness of the search and reduce potential language bias. Although in the final included studies there is no Persian study because of not having inclusion criteria, the investigation of Persian-language studies allowed for making sure about the existing of additional relevant evidence that may otherwise have been overlooked. All included studies were assessed using the same eligibility criteria to ensure consistency. However, studies published in other languages were not searched, which may still introduce some degree of language bias.

To address citation bias, we did not rely solely on highly cited articles or those published in high-impact journals; instead, we conducted a comprehensive and systematic search across multiple databases to capture both widely recognized and less frequently cited studies. Finally, to minimize multiple-publication bias, we carefully reviewed the included studies to identify and exclude duplicate reports or multiple publications from the same dataset, ensuring that each study contributed uniquely to the analysis.

Results

Overview

A total of 23 articles were included in this systematic review (Figure 1), comparing various wearable device technologies used in monitoring DFUs and their technical characteristics and clinical participant-related features (Tables 1 and 2).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart describing the selection of studies for the systematic review.

XSL•FO
RenderX

Table . Descriptive summary of articles included in the study (part 1).

No.	Authors, published year	Country	Study design	Sample size	Follow-up period	Foot ulcer Neuropathy	Type of diabetes	Application type	Monitored metrics	Interface
1	Hochlenert et al (2025) [28]	Germany	Clinical trial	17	7.5 months	Foot ulcer	Type 1 or 2	Bandage	Pressure, temperature, humidity, and steps	Smartwatch
2	Bulathsinghala et al (2024) [29]	Sri Lanka	Cohort	16	25 days	11 participants had foot ulcer	Type 1 or 2	Bandage	Blood flow and temperature	Smartphone
3	Cay et al (2024) [30]	United States	Clinical trial	62	3 months	Foot ulcer	Not reported	Boot	Offloading adherence, daily steps, and cadence	Smartwatch
4	Hu et al (2024) [31]	China	Clinical trial	23	No follow-up	None	Type 2	Anklet	Blood flow	Not reported
5	Park et al (2023) [32]	United States	Case series	14	No follow-up	None	None	Boot	Adherence, gait, and balance	Smartwatch
6	Tang et al (2023) [33]	United Kingdom	Case-control	6	No follow-up	None	Not reported	Insole	Pressure and shear stress	Not reported
7	Gupta et al (2023) [22]	India	Cross-sectional	50	No follow-up	Not reported	Not reported	Insole	Pressure	Not reported
8	Armstrong et al (2023) [34]	United States	Prospective study	10	3 months	Foot ulcer	Not reported	Boot	Pressure	Not reported
9	Khandakar et al (2022) [3]	Qatar	System analysis	12	No follow-up	Not reported	Not reported	Insole	Pressure and temperature	Smartphone
10	Reyzelman et al (2022) [35]	United States	Case-control	31	More than 50 days	Neuropathy	Type 1 or 2	Sock	Temperature	—[a]
11	Scholten et al (2022) [36]	United States	Cross-sectional	147	7 months	Neuropathy	Not reported	Sock	Temperature	—
12	Samarentsis et al (2022) [37]	Greece	Development evaluation	1	No follow-up	None	None	Insole	Pressure	PC[b]
13	Beach et al (2021) [38]	United Kingdom	Clinical trial	25	No follow-up	None	Type 1 or 2	Insole	Temperature	Smartphone
14	Du et al (2021) [39]	China	Clinical trial	6	6 months	Neuropathy	Type 1	Sensor	Gait and balance	Not reported
15	Torreblanca González et al (2021) [19]	Spain	Clinical trial	93	No follow-up	Not reported	Not reported	Sock	Temperature	Smartphone
16	Abbott et al (2019) [40]	United Kingdom	Clinical trial	58	18 months	Both	Type 1 or 2	Insole	Pressure	Smartwatch
17	Huchegowda et al (2019) [41]	India	Cross-sectional	30	No follow-up	Neuropathy	Type 2	Insole	Pressure	Smartphone

XSL•FO
RenderX

No.	Authors, published year	Country	Study design	Sample size	Follow-up period	Foot ulcer Neuropathy	Type of diabetes	Application type	Monitored metrics	Interface
18	Ming et al (2019) [20]	Germany	Clinical trial	300	24 months	Neuropathy and third group had history of foot ulcer	Type 1 or 2	Insole in slippers	Temperature	Smartphone
19	Reyzelman et al (2018) [42]	United States	Observational study	35	7 days	Neuropathy and the second and third groups include foot ulcers and pre-ulcer	Type 1 or 2	Sock	Temperature	Smartphone
20	Zhou et al (2018) [43]	United States	Cross-sectional	196	No follow-up	None	None	Anklet	Gait and balance	PC or tablet
21	Coates et al (2016) [44]	United Kingdom	Cross-sectional	16	No follow-up	None	None	Sensor	Acceleration, rotation, GSR[c], environmental temperature, humidity, force, skin temperature, and bioimpedance signals	Arduino devices
22	Grewal et al (2015) [45]	Qatar and United States	Clinical trial	39	1 month	Neuropathy	Type 2	Body-worn sensor technology	Balance, daily physical activity, triaxial accelerometer, gyroscope, and magnetometer	PC
23	Wrobel et al (2014) [46]	United States	Cross-sectional	27	No follow-up	Neuropathy	Type 1 or 2	Insole	Temperature, sudomotor function, gait, and balance	Not reported

[a]Not available

[b]PC: personal computer.

[c]GSR: galvanic skin response.

XSL•FO
RenderX

Table . Descriptive summary of studies included in the study (part 2).

Number	Authors (year)	Sensor location	Sensor count	Sensor type	Connectivity	Measurement range	Sampling and acquisition rates (Hz)
1	Hochlenert et al (2025) [28]	• Under the ulcer (variable)	1	• Custom built	Bluetooth	• Not reported	≈ 0.00167
2	Bulathsinghala et al (2024) [29]	• Left and right forefoot areas	2	• Not reported	Bluetooth	• ±2 °F	50
3	Cay et al (2024) [30]	• Shin	1 per leg	• Balance and gait IMU[a]	BLE[b]	• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	≈ 100
4	Hu et al (2024) [31]	• Medial forefoot • M5[c] • Heel • M1	4	• Laser doppler flowmetry	Bluetooth	• 0.8 - 4.5 mm/s	20
5	Park et al (2023) [32]	• Lower back+upper thighs • Shin	5	• Balance and gait IMU	BLE	• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	≈ 100
6	Tang et al (2023) [33]	• MH1[d] • MH5 • Hallux	4	• Custom built, capacitive	Wireless	• Pressure: 0 _ 300 kPa • Shear stress: ±90 kPa	100
7	Gupta et al (2023) [22]	• Lateral heel • Toes • Medial forefoot • Central forefoot • Lateral forefoot • Midfoot • Heel	126	• Piezoresistive	Wi-Fi	• 5 _ 900 kPa	40
8	Armstrong et al (2023) [34]	__[e]	—	• Compression pump	Not reported	—	—
9	Khandakar et al (2022) [3]	• Toes • Metatarsals • Medial and lateral arch • Heel	24	• Piezoelectric	BLE	• Temperature: 20 _ 50 ◦C • Pressure: 80 _ 600 kPa	40
10	Reyzelman et al (2022) [35]	• M1 • M3 • M5 • Hallux • Heel • Arch	6	• Resistive	—	• 20 _ 40 °C	≈ 0.1

XSL•FO
RenderX

Number	Authors (year)	Sensor location	Sensor count	Sensor type	Connectivity	Measurement range	Sampling and acquisition rates (Hz)
11	Scholten et al (2022) [36]	• M1 • M3 • M5 • Hallux • Heel • Arch	6	• Resistive	—	• 20 _ 40 °C	≈ 0.1
12	Samarentsis et al (2022) [37]	• Metatarsals • Toes • Lateral arch • Heel	16	• Capacitive	Bluetooth	• 41.5 _ 872.4 kPa	Not reported
13	Beach et al (2021) [38]	• MH1 • MH5 • Hallux • Heel	4	• Resistance	BLE	• 22 _ 42 ˚C	1
14	Du et al (2021) [39]	• Gait sensors • Bilateral lower shins • Balance sensors • Lower back • Lower left shin	4	• Balance and gait IMU	Bluetooth	• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	100
15	Torreblanca González et al (2021) [19]	• MH1 • MH5 • Heel • Medial and lateral arch • Hallux	6	• Resistive	Bluetooth	• 0 _ 43 ˚C	≈ 0.1
16	Abbott et al (2019) [40]	• MH1 • Lateral metatarsal heads • Hallux • Lateral toes • Lateral arch • Heel	8	• Resistive	Wireless	• 25 _ 225 mm Hg	≈ 8
17	Huchegowda et al (2019) [41]	• Not reported	Not reported	• Capacitive	Bluetooth	• Not reported	Not reported
18	Ming et al (2019) [20]	• MH1 • MH3 • MH5 • Hallux • Lateral • Heel	6	• Capacitive	BLE	• 15 _ 600 kPa	2
19	Reyzelman et al (2018) [42]	• M1 • M3 • M5 • Hallux • Midfoot • Heel	6	• Resistive	Bluetooth	• 20 _ 40 °C	≈ 0.1
20	Zhou et al (2018) [43]		2		Bluetooth		≈ 100

XSL•FO
RenderX

Number	Authors (year)	Sensor location	Sensor count	Sensor type	Connectivity	Measurement range	Sampling and acquisition rates (Hz)
		• Left and right lower shins		• Balance and gait IMU		• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	
21	Coates et al (2016) [44]	• M1 • M5 • Heel • Hallux	42	• Accelerometer: IMU • Humidity: capacitive • Temperature: not reported • GSR:[f] IMU • Bioimpedance: digital frequency synthesis • Force: piezoresistive	Bluetooth	• Accelerometer: ±16 g • Rotation: ±2000 ◦s−1 • Humidity: 0% - 99 % RH • Temperature: -40 - 125 ◦C • GSR: a 0 - 5000 kΩ • Bioimpedance: 0 - 1023 AU[g] • Force: 0 - 140 N • Temperature skin: 20 - 40 ◦C	20
22	Grewal et al (2015) [45]	• Shin • Thigh • Lower back	5	• Balance and gait IMU	Bluetooth	• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	≈ 100
23	Wrobel et al (2014) [46]	• Midfoot • Forefoot • Hindfoot • Entire foot	≈ 5	• Balance and gait IMU	Bluetooth	• Accelerometer: ±2 g • Gyroscope: ±2000 deg/s	≈ 100

[a]IMU: internal measurement unit.

[b]BLE: Bluetooth low energy.

[c]M: metatarsal.

[d]MH: metatarsal head.

[e]Not applicable.

[f]GSR: galvanic skin response.

[g]AU: arbitrary units.

Interrater reliability between the 2 reviewers was assessed using Cohen kappa. The level of agreement was high (κ=0.87; Table S2 in Multimedia Appendix 1).

Quality Assessment

Most of the included studies were good in terms of quality. Of the 23 included papers, 4 [40,43-45] studies met all MMAT criteria (2 RCTs and 2 nonrandomized controlled trials). In 7 [19,20,28,30,31,38,39] RCTs, assessors blinded to the intervention were not clarified (Table S1 in Multimedia Appendix 1).

Year-Oriented Analysis

The dataset includes studies published between 2014 and 2025, highlighting a growing academic interest in the field. The distribution of publications indicates an upward trend, with the highest number of studies published in 2022 (n=4) [3,35-37] and 2023 (n=4) [22,32-34], with 2024 and 2019 each n=3 [20,29-31,40,41]. In contrast, the fewest publications were recorded in 2014, 2015, 2016, and 2025, with each year contributing only one study [28,44-46]. A notable gap in research output was observed between 2016 and 2018, during which no studies met the inclusion criteria. In 2020, 2 [47,69] studies used mats; however, since mats are not considered wearable devices, these studies were excluded.

XSL•FO
RenderX

Country-Based Analysis

The geographical distribution of the reviewed studies shows a strong concentration of research in the United States, where 9 [30,32,34-36,42,43,45,46] studies were conducted. Additionally, 4 studies were conducted in the United Kingdom [33,38,40,44], while China [31,39], Qatar [3,45], India [22,41], and Germany [20,28] each contributed 2 studies. Other studies were conducted in Greece [37], Spain [19], and Sri Lanka [29].

Study Design Analysis

The reviewed studies used a variety of research designs. Nine [19,20,28,30,31,38-40,45] were identified as clinical trials, while 6 [22,36,41,43,44,46] studies used a cross-sectional design. Two [33,35] studies used a case-control design, and 2 [3,37] used system analysis and development evaluation. One observational [42], one prospective study [34], one case series [32], and one cohort study [29] were also identified.

Sample Size Analysis

The sample sizes across the reviewed studies showed considerable variation, ranging from as few as a single participant [37] to as many as 300 participants [20]. Despite this wide range, the average sample size was approximately 26 participants.

Follow-Up Period Analysis

The study with the longest follow-up period tracked participants for 24 months [20]. This clinical trial aimed to assess the effectiveness of telemedicine and foot temperature monitoring in reducing the risk of ulcer formation. Several other studies followed participants for shorter durations, specifically 18 [40], 7 [28,36], 6 [39], and 3 months [30,34]. However, most of the studies did not include a follow-up period and relied on a single laboratory test [3,19,22,31-33,37,38,41,43,44,46].

Participant Analysis

Nine [20,35,36,39-42,45,46] studies included participants with a history of neuropathy, while 7 [20,28-30,34,40,42] focused on those with either active ulcers or a history of DFUs. The remaining studies either did not specify participants' neuropathy or ulcer history or confirmed that participants had no such conditions. Eight [20,28,29,35,38,40,42,46] studies included participants with a history of diabetes, either type 1 or 2. Three [31,41,45] studies specifically involved participants with type 2 diabetes, and one [39] study included participants with type 1 diabetes. The remaining studies did not specify the type of diabetes.

Wearable Device Applications Analysis

Nine [3,20,22,33,37,38,40,41,46] studies developed insoles, while 4 [19,35,36,42] focused on socks, and 3 [30,32,34] on boots. The remaining studies either did not specify the type of application or involved only sensors and bandages.

Key Metrics Analysis

Eleven [3,19,20,28,29,35,36,38,42,44,46] studies reviewed focused on monitoring temperature. Eight [3,22,33,34,37,40,41,44] studies monitored pressure. In addition to temperature and pressure, several other metrics were considered important for ulcer prediction. Gait and balance were reported in 4 [32,39,43,46] studies, humidity in 2 [28,44] studies, daily step count in 2 [28,30] studies, and blood flow in 2 [29,31] studies. Force [44], shear stress [33], galvanic skin response [44], and bioimpedance signals [44] were each reported in one study.

User Interface Analysis

Smartphones were used in 7 [3,19,20,29,38,41,42] of the reviewed studies, likely due to their portability and ease of integration with wearable sensor systems. Their built-in computing power and wireless connectivity enabled real-time data processing and supported user-friendly application interfaces.

Personal computers were used in 3 [37,43,45] studies that required more complex computational analysis. These tasks included machine learning–based gait assessment and high-resolution pressure mapping, where greater processing capacity was required to handle large datasets and detailed modeling.

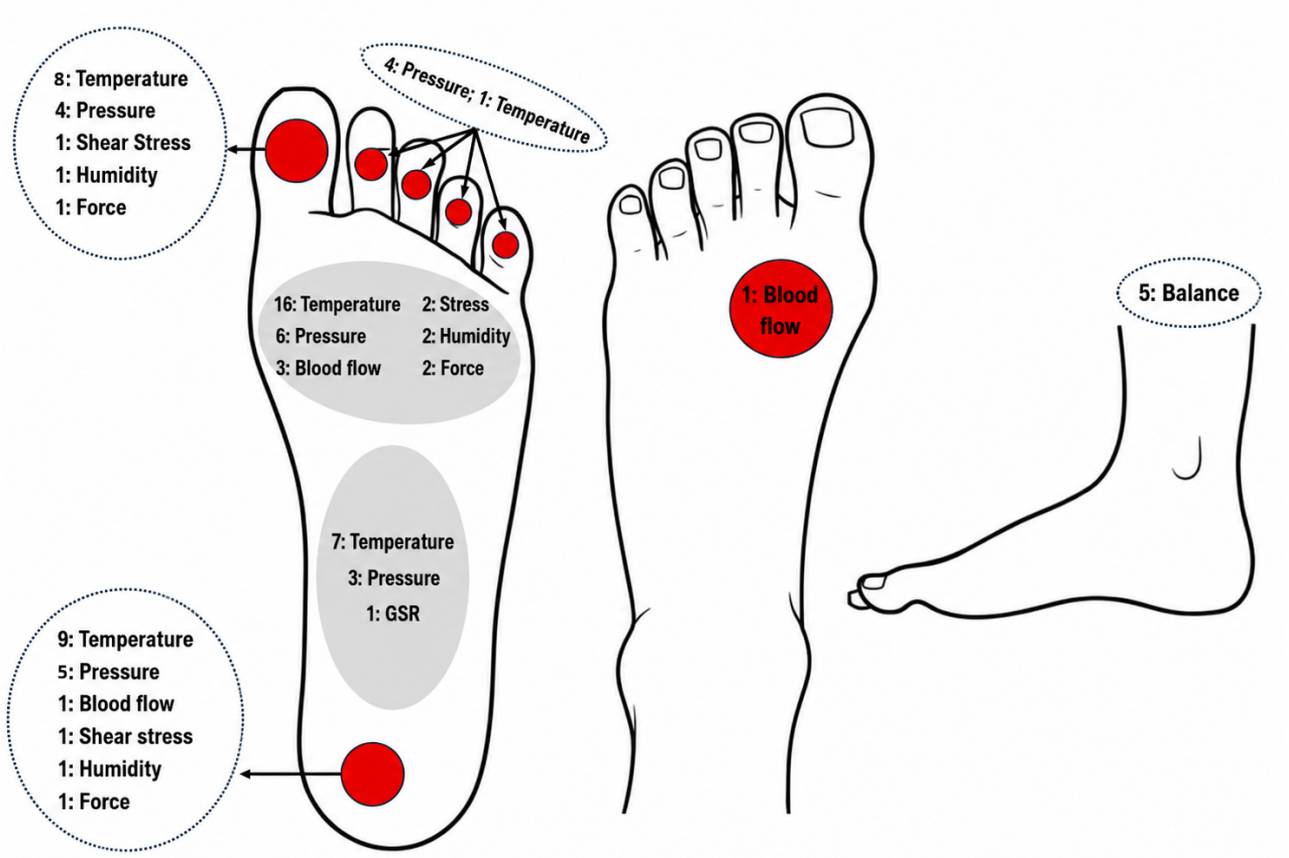
Technical or Engineering Outcomes

Sensors' Location Analysis

The studies used various sensor placement strategies to capture foot-related biomechanical data. Most studies measuring temperature placed sensors on the first, third, and fourth metatarsals, as well as on the heel, arch (medial and lateral), and hallux [19,20,35,36,38,42,44,46]. For pressure measurements, sensors were typically placed on the first metatarsal head, lateral metatarsal heads, hallux, lateral toes, lateral foot, toes, arch (medial and lateral), and heel [22,33,37,40]. Moreover, studies focusing on gait and balance commonly placed sensors on the lower back and dominant lower shin [30,32,39,43-46] (Figure 2).

XSL•FO
RenderX

Figure 2. Distribution of studies by sensor location and number of sensors. GSR: galvanic skin response.



Number of Sensors Analysis

The number of sensors used across the reviewed studies varied considerably, reflecting differences in study objectives, measurement complexity, and technological capabilities. Sensor configurations ranged from single-sensor systems to arrays consisting of 5 [32,45,46], 6 [19,20,35,36,42], 8 [40], 16 [37], 24 [3], or even 42 sensors [44] (Figure 2).

Sensors' Technology Analysis

Various sensor technologies were used, including piezoresistive, capacitive, and piezoelectric sensors. Seven [30,32,39,43-46] studies used a balance and gait internal measurement unit, 6 [19,35,36,38,40,42] studies referenced resistive sensors, 5 [20,33,37,41,44] studies mentioned capacitive sensors, and 3 [3,22,44] studies mentioned using piezoresistive and piezoelectric sensors. The remaining studies used laser Doppler flowmetry [31] or a compression pump [34], while others did not report the sensor technology used in developing their wearable devices.

Communication Method Analysis

Generally, 2 main types of communication protocols are used in microcontroller systems: one facilitates communication between the sensors and the microcontroller (eg, Interintegrated Circuit, Serial Peripheral Interface, and Universal Asynchronous Receiver-Transmitter), and the other manages communication between the microcontroller and the reference device or server (eg, Bluetooth and Wi-Fi). The reviewed studies primarily used wireless communication technologies, with Bluetooth and Wi-Fi being the most common. Among the 23 studies analyzed, Bluetooth was the preferred method, used in 12 [19,28,29,31,37,39,41-46] studies. Notably, 5 [3,20,30,32,38] studies specified the use of Bluetooth low energy (BLE), emphasizing energy-efficient data transmission, which is particularly beneficial for wearable sensor applications. In addition to Bluetooth, other general wireless communication technologies were used in 3 studies, although the specific protocols were not always clearly defined [22,33,40].

Range of Measurement Analysis

The measurement range varied considerably across the reviewed studies, primarily depending on the type of sensor and its intended application. Specifically, pressure sensors demonstrated a broad detection capability, spanning from 5 kPa to 900 kPa [3,20,22,33,37]. This wide range allows for the assessment of both low-pressure and high-impact forces, making such sensors suitable for applications in gait analysis, foot biomechanics, and clinical diagnostics.

Similarly, temperature sensors were used in several studies, with a reported measurement range of 0 °C to 50 °C [3,19,29,35,36,38,42,44]. This range aligns with physiological temperature variations and external environmental influences, both of which are crucial for monitoring foot health, particularly in patients with diabetes and individuals with circulatory disorders.

XSL•FO
RenderX

Acquisition Rate or Sampling Frequency

The sampling frequency of sensor systems varied across the reviewed studies, ranging from 0.001 Hz to 100 Hz. The reported sampling frequency may not reflect the maximum capability of the sensors. Typically, we reported the value presented in the articles; however, it is possible that the sensors can record or acquire data at higher frequencies.

Higher sampling rates are generally associated with improved temporal resolution, enabling more precise capture of dynamic biomechanical events. The highest reported acquisition rate was 100 Hz [30,32,33,39,43,45,46], which facilitates detailed gait analysis by capturing rapid changes in foot pressure and movement patterns. However, higher rates also increase the risk of recording noise [70].

Lower sampling frequencies (eg, 1 Hz) may still be adequate for applications where slower physiological changes are monitored, such as prolonged pressure distribution assessments. However, for real-time gait analysis or high-impact activities, higher acquisition rates are preferable to avoid data loss and improve the accuracy of motion tracking.

Overall, the findings suggest that most current wearable systems are mainly designed for preventive monitoring, particularly through measuring plantar pressure, temperature, and gait patterns. However, considerable variation in sensor choice, communication methods, and reporting standards makes it difficult to directly compare devices across studies. More consistent technical reporting and standardized benchmarking approaches are needed to enable clearer comparisons and support future clinical implementation.

Clinical Outcomes

Wearable sensors for DFUs in patients with diabetes have demonstrated a broad range of positive clinical outcomes across multiple studies (Table S3 in Multimedia Appendix 1).

These technologies enabled early detection and prevention of foot ulcers and related complications, facilitating timely intervention and treatment [3,19,29,31,35,39,40,42,43,46].

A small subset of studies focused on patients with active DFUs, where wearable technologies were associated with accelerated wound healing, increased wound closure rates, reduction in wound area, and pain relief. For example, 2 [28,34] studies reported that wearable sensors accelerated wound healing and increased wound closure rates, while others showed significant wound area reduction and pain relief [34,41].

User-Level Outcomes

Wearable sensors demonstrated substantial benefits in the early detection and prevention of DFUs across multiple studies. These technologies enabled timely intervention by continuously monitoring risk factors such as plantar pressure and temperature. In 7 [28,30,35,37,40,42,45] studies, these devices supported patient-centered care by empowering patients to self-monitor, improving engagement, and enabling personalized treatment plans tailored to individual risk factors and healing barriers. Real-time tracking of patients' adherence to offloading allowed proactive intervention when patients deviated from prescribed care [32]. Remote and at-home monitoring, reported in 5 studies, reduced the need for frequent clinic or hospital visits, making long-term management more accessible [3,19,30,31,35]. Functional improvements such as better gait, balance, postural stability, and overall mobility were described in 5 [32,33,39,41,45] studies. The cost-effectiveness of these technologies was highlighted in 3 [3,29,37] studies. Mental health benefits, including increased confidence and reduced anxiety, were reported in one study [45], while quality of life improvements were noted in another [28]. Furthermore, wearable sensors proved suitable for routine clinical use and long-term monitoring without disrupting patients' daily activities [19,31].

Discussion

Overview of Research Trends and Study Characteristics

As the results indicate, there was no publication between 2016 and 2018 and low publication before 2016; however, the rise in studies since 2019 demonstrates a renewed research interest, likely driven by advances in microcontrollers and miniaturized sensors. Additionally, inclusion studies were conducted in various countries, from the United States to China. This pattern may reflect a growing global recognition of the importance of diabetic foot monitoring. The geographical diversity could indicate increasing international awareness; however, it may be influencing the adoption and development of these technologies [71]. Furthermore, our findings indicate that the majority of the included studies were clinical trials. RCTs ensure the production of reliable and robust data, which is essential for regulatory approvals, clinical guideline development, and public health policy decisions [72].

Since 2019, growing international research—primarily RCTs—reflects renewed interest and robust evidence supporting the adoption of wearable technologies for patients with DFUs. The prevalence of RCTs in this review may reflect an emphasis on generating high-quality evidence to support the use of wearable technologies for diabetic foot monitoring.

Although many of the 23 reviewed studies were RCTs—which strengthens their internal validity—this design does not automatically address 2 critical methodological limitations, including sample size and follow-up duration. The mean sample size across the included studies was only 26 participants, and many studies relied on single-session laboratory tests without any follow-up. This pattern indicates that most studies in this field were underpowered for developing generalizable prediction models, increasing the risk of overfitting and reducing external validity. Furthermore, the lack of follow-up means that dynamic preulcerative changes cannot be captured. Consequently, even within RCT designs, most studies provide only a biomechanical snapshot rather than a valid prognostic assessment, a limitation that some of the included studies explicitly acknowledged [28,31].

Technical and Clinical Features of Wearable Systems

Sensor placement strategies across studies exhibited considerable heterogeneity, with common locations including

XSL•FO
RenderX

the metatarsal heads, hallux, heel, midfoot, and arch (medial and lateral). These placements align with established biomechanical principles, as they enable the capture of temperature, pressure distribution, load transfer, and gait characteristics [73].

The multiple sensor locations on the foot are essential for effective detection and prediction of DFUs, as highlighted by multiple studies in this review. This strategic placement enables comprehensive monitoring of the key areas where ulcers are most likely to develop, improving early identification of risk and timely intervention. Jones et al [74] indicate that integrating parameters such as temperature, pressure, shear stress, and humidity across anatomically important sites yields richer data for prognosis. Many current models focusing on single parameters demonstrate limited predictive specificity (<50%), underscoring the need for multipoint and multisensor arrays precisely positioned to capture the foot microclimate comprehensively [74]. Although higher sensor density may improve spatial resolution, it remains unclear from the included studies whether this translates into better clinical prediction of DFUs, as no included study directly compared different sensor densities within the same population.

Despite these biomechanical rationales, the present review identified considerable heterogeneity in sensor placement protocols across the 23 included studies. This heterogeneity represents a key methodological limitation emerging from the reviewed studies. While the lack of a universal standardized protocol allows flexibility to address different research questions, it substantially complicates direct comparisons between studies and limits the feasibility of meta-analytic synthesis. Future research should therefore work toward establishing consensus-based guidelines for sensor placement to improve comparability, reproducibility, and clinical translatability across studies.

Also, a variety of foot monitoring systems were designed based on clinical and research requirements to address various needs [75]. Smart insoles appeared to be the most commonly developed solution, possibly due to their effectiveness. Socks and boot-based monitoring systems are other devices that may offer better support and stable sensor placement at high-risk anatomical sites [32,76]. The remaining studies either did not clearly specify the type of system or mainly worked with basic sensors and bandages. The range of application types shows the need for flexible solutions that can fit different users, such as people walking normally and engaging in sports, patients in clinics, or those needing home monitoring. Future research should continue to improve these wearable systems, making them more comfortable, easier to use, and better at collecting reliable data to help in real-world health care settings [29].

The reviewed studies consistently emphasize the importance of participant clinical history related to neuropathy and foot ulceration, particularly focusing on individuals with active DFUs or due to the high risk of recurrence. This participant characteristic is crucial as it identifies populations at high risk for ulcer recurrence and serious complications. Variation in participant selection across studies mirrors differing research goals. Studies aiming for early detection of risk factors or initial neuropathy may enroll broader diabetic populations without active ulcers to develop screening tools [77]. In contrast, studies evaluating monitoring technologies or treatment effectiveness often focus on patients with active or healed ulcers to assess prevention of recurrence and ulcer healing outcomes [78].

The number of sensors used across the 23 reviewed studies varied considerably, ranging from single-sensor systems to arrays of up to 42 sensors. Configurations included 5, 6, 8, 16, 24, and 42 sensors, depending on study objectives and measurement complexity. Studies requiring detailed localized pressure and temperature data tended to use higher sensor densities, whereas those focusing on broader gait parameters used fewer sensors for practicality.

External reviews reported that increased sensor density enhances spatial resolution, potentially boosting the precision of pressure and movement monitoring. Different sensor approaches for diabetic foot monitoring discuss trade-offs between sensor density, imaging modalities, and integration with AI tools for remote management. The sensor type and data resolution must align with application contexts and technical constraints [79]. Multifaceted wearable systems use a limited number of sensors embedded in insoles to capture clinically relevant pressure and temperature data while balancing wearability and adherence in daily life [80]. Reviews of plantar pressure and temperature sensing technologies confirm that high-density mats provide detailed foot surface mapping but face limitations for use in real-world ambulatory settings, where moderate sensor counts suffice and improve user compliance [81,82].

In this review, studies with higher sensor counts (eg, 42 sensors) were primarily technical in nature and did not report direct clinical outcomes such as ulcer prevention or healing. In contrast, studies using 5‑6 sensors (typically combining temperature and pressure) more frequently reported preventive outcomes. This pattern suggests that increasing the number of sensors does not necessarily translate into improved clinical outcomes, and there may be an optimal threshold beyond which additional technical complexity adds little clinical value.

Furthermore, no included study directly compared different sensor densities within the same population to determine whether higher spatial resolution improves clinical prediction of DFUs. The included studies also did not provide data on the trade-off between sensor density and patient adherence or comfort. Based on the available evidence, moderate sensor configurations (eg, 5‑8 sensors) may be sufficient for routine clinical use, but this hypothesis requires direct testing in future research.

In terms of sensing technologies used for plantar pressure monitoring, three main types of sensors are commonly used: piezoelectric, piezoresistive, and capacitive. Each offers distinct advantages in terms of sensitivity, durability, and energy efficiency. Despite the technological diversity available, most of the studies included in this review did not report the type of sensor technology used in their devices and typically name the commercial system.

This lack of detail may limit the ability to evaluate device performance, compare outcomes across studies, or draw

XSL•FO
RenderX

conclusions about the relative effectiveness of different sensor types in predicting DFUs. Capacitive and resistive sensors dominate recent in-shoe pressure measurement, but no current system fully meets all ideal criteria—like accuracy over multiple load cycles, comfort, wireless communication, and affordability—highlighting ongoing challenges in sensor development. Furthermore, variability in measurement outcomes across devices complicates the assessment of their relative efficacy, underscoring the need for transparency in reporting sensor technologies [83,84].

Variation in communication methods reflects trade-offs between power consumption, data transfer rates, and connectivity stability. BLE has increasingly become the protocol of choice in numerous Internet of Things (IoT) devices, which commonly implement lightweight authentication and encryption methods to ensure secure communication [85]. The preference for wireless communication underscores the need for mobility and real-time data acquisition in gait and pressure analysis. Wi-Fi, though less frequently mentioned, offers advantages in terms of data transfer speed and network connectivity, making it a viable option for applications requiring continuous high-throughput data transmission. The problem is that Wi-Fi could be vulnerable to cyber-attacks. Wireless electronic sensors have the potential to reduce health care expenses by allowing physicians to remotely track key physiological information [86].

Wearable systems for diabetic foot monitoring use diverse sensor placements and technologies to capture critical biomechanical and physiological data. Multipoint sensor arrays improve early detection of DFUs by monitoring pressure, temperature, and other parameters in high-risk areas. Smart insoles and wireless communication, particularly via BLE, enhance usability and real-time remote monitoring capabilities. Despite these advancements, challenges remain regarding sensor accuracy, wearability, and the transparent reporting of technologies. Future improvements should target user comfort, data reliability, and integration with advanced analytics for better clinical outcomes.

Key Metrics and Clinical Outcomes

Participant selection across the reviewed studies mainly focused on the importance of diabetic foot monitoring. Most studies included participants with a history of diabetes, either as the main group or as part of a mixed sample. Including people with diabetes highlights the practical use of sensor-based monitoring systems for the early detection and prevention of DFUs.

The evidence shows that most wearable or insole-based pressure sensor studies include patients with diabetes, reflecting a practical research approach targeting those who will most benefit from early detection of increased plantar pressure or temperature changes [87]. Large RCTs assessing the effectiveness of these monitoring systems also recruit diabetic cohorts predominantly to maximize clinical relevance and event detection [88].

Temperature and pressure are reported as key parameters in reviewed studies. The evidence also indicates that temperature and pressure are important early indicators, highlighting the significant role these 2 factors play in predicting DFUs. External evidence suggests temperature is a sensitive early warning sign, though models relying on a single parameter rarely exceed 50% specificity [74].

Patients with diabetes have higher plantar pressure than healthy individuals. Along with pressure and temperature, other factors like gait, balance, and skin humidity were monitored, as they influence DFU risk. Daily step counts provided insights into movement patterns and stress on the foot. Additional parameters—such as blood flow, shear stress, and galvanic skin response—were explored to offer a more comprehensive assessment of foot health and ulcer risk. Additional parameters such as gait, balance, humidity, blood flow, and shear stress were explored in a subset of studies [78,83]. Building on this, wearable sensors also improve clinical outcomes and patient care in diabetic foot management. Wearable sensors for DFUs in patients with diabetes provide a broad spectrum of clinical benefits, including early detection, prevention, improved healing, patient empowerment, remote monitoring, functional improvements, and cost-effectiveness. These outcomes support the integration of wearable technology into routine diabetic foot care to enhance patient outcomes and optimize health care resource use. While a few studies did not report direct clinical outcomes, the overall evidence strongly supports the integration of wearable sensor technology into routine DFU management to improve clinical outcomes.

Based on the included studies, wearable sensors demonstrated clinical benefits, including early detection, prevention, improved healing, patient empowerment, remote monitoring, functional improvements, and cost-effectiveness. External studies have also reported reduced neuropathic pain [89]; decreased acute care use, including hospitalizations and emergency visits, in a substantial proportion of studies [90]; and improved mental health outcomes [91], although these involved different device types and populations.

However, 3 important distinctions should be noted: (1) the external studies used different device types (eg, nerve stimulation devices, telehealth programs) rather than the in-shoe pressure/temperature sensors reviewed here; (2) the external studies often had larger sample sizes (eg, n=550 in [89]) compared to the mean of 26 participants in the present review; and (3) the external studies focused on outcomes such as pain and mental health, whereas the present review found that most included studies focused on biomechanical parameters (pressure, temperature) rather than direct patient-reported outcomes. Therefore, direct extrapolation of these external findings to wearable in-shoe monitoring systems for DFU prevention should be made with caution.

Also, these technologies provide comprehensive foot health assessment by integrating multiple physiological measures, advancing personalized care. Wearable devices also demonstrate significant clinical benefits, including neuropathic pain reduction, improved mobility, and decreased ulcer incidence. Remote patient monitoring and telehealth further enhance clinical outcomes and mental health by enabling timely, accessible support.

Overall, these related studies support and extend the initial findings by suggesting that wearable sensors may have potential benefits in the prevention and management of DFUs.

XSL•FO
RenderX

The integration of wearable monitoring devices in diabetic foot care has shown substantial promise in improving early detection, prevention, and management of foot complications. Wearable technologies have emerged as promising tools for enhancing foot health monitoring in individuals at risk and those who had DFUs previously.

This review has several limitations that should be considered when interpreting the findings. First, there was heterogeneity across the included studies in terms of study design, sample size, patient populations, device types, and outcome measures, which limited the ability to directly compare results or draw generalizable conclusions.

Second, many studies lacked comprehensive technical reporting, including details on sensor calibration, validation against gold standards, and data processing methods. This variability in reporting limits the reproducibility and comparability of findings across studies and, in some instances, necessitated reliance on external evidence to contextualize or support certain technical interpretations.

Finally, few studies reported strong clinical outcomes, such as ulcer occurrence, recurrence, infection, or amputation. Most studies relied on indirect measures, such as pressure or temperature, which may not fully represent real clinical benefit.

Future Directions in Wearable Technology

The integration of AI and multimodal sensing represents a promising direction for the evolution of wearable technologies in DFU management. Multimodal sensing combines parameters such as plantar pressure, temperature, and shear stress. Emerging devices, including smart insoles, sensor-embedded socks, and external foot sensors, enable continuous monitoring of multiple physiological and biomechanical parameters, such as pressure, temperature, humidity, blood flow, gait, and balance. This shift toward data-driven, multiparameter systems may enhance predictive performance and improve clinical trust, addressing key limitations associated with single-parameter monitoring approaches.

Clinical Implications and Patient-Centered Care

By capturing diverse data points, wearable technologies offer a comprehensive and multifactorial approach to assessing DFU risk. Their integration into routine diabetic care has the potential to support real-time monitoring and personalized feedback, enabling earlier detection of abnormalities and timely intervention.

This technology-driven approach may enhance patient empowerment and self-management, leading to improved adherence and clinical outcomes. Furthermore, continuous data collection can support more informed clinical decision-making and facilitate the development of personalized care strategies tailored to individual patient needs.

Conclusion

Finally, few studies reported robust clinical outcomes, such as ulcer occurrence, recurrence, infection, or amputation. Most studies relied on indirect measures, such as pressure or temperature, which may not fully reflect real clinical effectiveness. Future research should focus on developing standardized reporting frameworks for wearable technologies to enable more consistent and meaningful comparisons across studies. There is also a need for pragmatic, real-world trials that include adherence-based interventions to better understand long-term outcomes and real-world performance. In addition, further high-quality studies are needed to evaluate the clinical benefits of these technologies and clarify their role within existing care pathways. If supported by stronger evidence, integrating wearable technologies into routine diabetic foot care may help improve adoption and patient outcomes.

Acknowledgments

The authors would like to thank Prof Mohammad Hossein Mehrolhasani for his general academic support during this research. Artificial intelligence (AI) was used as a language-editing service during the writing process of the study to check grammar and spelling. AI was not used to perform core research tasks, such as generating scientific insights, analyzing and interpreting data, and drawing conclusions.

Funding

The authors declared no financial support was received for this work.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: HE, SS
Data curation: HE, SS
Formal analysis: HE
Investigation: HE
Methodology: HE, SS
Project administration: HE, SS, SSE, DR
Resources: HE, SS

XSL•FO
RenderX

Supervision: SSE, HS, AS, DR
Validation: HE, SS, DR
Visualization: HE, SS
Writing – original draft: HE, SS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality assessment, interrater agreement, and clinical outcomes of the included studies.
[DOCX File, 107 KB - diabetes_v11i1e84463_app1.docx]

Checklist 1
PRISMA checklist.
[PDF File, 67 KB - diabetes_v11i1e84463_app2.pdf]

References

1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
2. McDermott K, Fang M, Boulton AJM, Selvin E, Hicks CW. Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers. Diabetes Care 2023 Jan 1;46(1):209-221. [doi: 10.2337/dci22-0043] [Medline: 36548709]
3. Khandakar A, Mahmud S, Chowdhury MEH, et al. Design and implementation of a smart insole system to measure plantar pressure and temperature. Sensors (Basel) 2022 Oct 7;22(19):7599. [doi: 10.3390/s22197599] [Medline: 36236697]
4. Uçkay I, Aragón-Sánchez J, Lew D, Lipsky BA. Diabetic foot infections: what have we learned in the last 30 years? Int J Infect Dis 2015 Nov;40:81-91. [doi: 10.1016/j.ijid.2015.09.023] [Medline: 26460089]
5. Eleftheriadou I, Kokkinos A, Liatis S. Atlas of the Diabetic Foot: John Wiley & Sons; 2019:1-26.
6. Whicher CA, O'Neill S, Holt RIG. Diabetes in the UK: 2019. Diabet Med 2020 Feb;37(2):242-247. [doi: 10.1111/dme.14225] [Medline: 31901175]
7. Kerr M, Rayman G, Jeffcoate WJ. Cost of diabetic foot disease to the National Health Service in England. Diabet Med 2014 Dec;31(12):1498-1504. [doi: 10.1111/dme.12545] [Medline: 24984759]
8. Leese GP, Stang D, Pearson DW, Scottish Diabetes Foot Action Group. A national approach to diabetes foot risk stratification and foot care. Scott Med J 2011 Aug;56(3):151-155. [doi: 10.1258/smj.2011.011113] [Medline: 21873720]
9. Mackenzie P. Diabetes Footcare Pathway Blueprint (2017). : NHS England; 2017 URL: https://www.england.nhs.uk/north/wp-content/uploads/sites/5/2018/05/NWCSN_Diabetes_Footcare_Final_Report_2017-1.pdf [accessed 2024-07-19]
10. Lazarou I, Fiska V, Mpaltadoros L, et al. Stepping forward: a scoping systematic literature review on the health outcomes of smart sensor technologies for diabetic foot ulcers. Sensors (Basel) 2024 Mar 21;24(6):2009. [doi: 10.3390/s24062009] [Medline: 38544271]
11. Lim JZM, Ng NSL, Thomas C. Prevention and treatment of diabetic foot ulcers. J R Soc Med 2017 Mar;110(3):104-109. [doi: 10.1177/0141076816688346] [Medline: 28116957]
12. Schaper NC, van Netten JJ, Apelqvist J, et al. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3657. [doi: 10.1002/dmrr.3657] [Medline: 37243927]
13. Bakker K, Apelqvist J, Schaper NC, IWGotDFE B. Practical guidelines on the management and prevention of the diabetic foot 2011. Diabetes Metabolism Res 2012 Feb;28(S1):225-231 [FREE Full text] [doi: 10.1002/dmrr.2253]
14. Crawford F, Nicolson DJ, Amanna AE, Smith M. Reliability of the evidence to guide decision-making in foot ulcer prevention in diabetes: an overview of systematic reviews. BMC Med Res Methodol 2022 Oct 20;22(1):274. [doi: 10.1186/s12874-022-01738-y] [Medline: 36266628]
15. Roy S, Nagabooshanam S, Chauhan N, Kumar R, Wadhwa S, Mathur A. Design and development of a novel flexible molecularly imprinted electroanalytical sensor for the monitoring of diabetic foot ulcers. Surf Interfaces 2021 Oct;26:101310. [doi: 10.1016/j.surfin.2021.101310]
16. Tehrani K, Michael A. Introduction to wearable healthcare technology. : University of Jyväskylä; 2017. [doi: 10.13140/RG.2.2.28570.90564]
17. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph beddit sleep tracker: a validation study. J Clin Sleep Med 2019 Mar 15;15(3):483-487. [doi: 10.5664/jcsm.7682] [Medline: 30853052]

XSL•FO
RenderX

18. Bint A, Mali HS, Malik A, Rosenthal J. Contemporary technologies in wearable orthoses and their potential intervention for clubfoot treatment: a review. Proc Inst Mech Eng H 2025 Oct;239(10):979-999. [doi: 10.1177/09544119251387177] [Medline: 41146528]
19. Torreblanca González J, Gómez-Martín B, Hernández Encinas A, Martín-Vaquero J, Queiruga-Dios A, Martínez-Nova A. The use of infrared thermography to develop and assess a wearable sock and monitor foot temperature in diabetic subjects. Sensors (Basel) 2021 Mar 5;21(5):1821. [doi: 10.3390/s21051821] [Medline: 33807804]
20. Ming A, Walter I, Alhajjar A, Leuckert M, Mertens PR. Study protocol for a randomized controlled trial to test for preventive effects of diabetic foot ulceration by telemedicine that includes sensor-equipped insoles combined with photo documentation. Trials 2019 Aug 22;20(1):521. [doi: 10.1186/s13063-019-3623-x] [Medline: 31439007]
21. Schaper NC, van Netten JJ, Apelqvist J, et al. Practical guidelines on the prevention and management of diabetic foot disease (IWGDF 2019 update). Diabetes Metab Res Rev 2020 Mar;36 Suppl 1(S1):e3266. [doi: 10.1002/dmrr.3266] [Medline: 32176447]
22. Gupta S, Jayaraman R, Sidhu S, et al. Diabot: development of a diabetic foot pressure tracking device. J 2023;6(1):32-47. [doi: 10.3390/j6010003]
23. Srass H, Ead JK, Armstrong DG. Adherence and the diabetic foot: high tech meets high touch? Sensors (Basel) 2023 Aug 3;23(15):6898. [doi: 10.3390/s23156898] [Medline: 37571682]
24. Srivastava T, Thiagarajan G, Alon US, et al. Role of biomechanical forces in hyperfiltration-mediated glomerular injury in congenital anomalies of the kidney and urinary tract. Nephrol Dial Transplant 2017 May 1;32(5):759-765. [doi: 10.1093/ndt/gfw430] [Medline: 28339567]
25. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
26. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
27. Pace R, Pluye P, Bartlett G, et al. Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
28. Hochlenert D, Bogoclu C, Cremanns K, et al. Sensor-assisted wound therapy in plantar diabetic foot ulcer treatment: a randomized clinical trial. J Diabetes Sci Technol 2025 May;19(3):692-698. [doi: 10.1177/19322968231213095] [Medline: 38006228]
29. Bulathsinghala RL, Wijeyaratne SM, Fernando S, Jayawardana TSS, Senadhipathi Mudiyanselage VUI, Kankanamalage SLS. Wearable medical apparatus to monitor temperature and pulsatile-blood-flow signal on forefoot to predict diabetic foot ulcers. Res J Text Apparel 2024 Apr 10;28(2):169-184. [doi: 10.1108/RJTA-11-2021-0135]
30. Cay G, Finco MG, Garcia J, McNitt-Gray JL, Armstrong DG, Najafi B. Towards a remote patient monitoring platform for comprehensive risk evaluations for people with diabetic foot ulcers. Sensors (Basel) 2024 May 8;24(10):2979. [doi: 10.3390/s24102979] [Medline: 38793835]
31. Hu XX, Xing XM, Zhang ZM, et al. Wearable laser doppler flowmetry for non-invasive assessment of diabetic foot microcirculation: methodological considerations and clinical implications. J Biomed Opt 2024 Jun;29(6):065001. [doi: 10.1117/1.JBO.29.6.065001] [Medline: 38737791]
32. Park C, Mishra R, Vigano D, et al. Smart offloading boot system for remote patient monitoring: toward adherence reinforcement and proper physical activity prescription for diabetic foot ulcer patients. J Diabetes Sci Technol 2023 Jan;17(1):42-51. [doi: 10.1177/19322968211070850] [Medline: 35048739]
33. Tang J, Bader DL, Moser D, et al. A wearable insole system to measure plantar pressure and shear for people with diabetes. Sensors (Basel) 2023 Mar 15;23(6):3126. [doi: 10.3390/s23063126] [Medline: 36991838]
34. Armstrong DG, Orgill DP, Glat PM, et al. A single arm prospective feasibility study evaluating wound closure with a unique wearable device that provides intermittent plantar compression and offloading in the treatment of non-healing diabetic foot ulcers. Int Wound J 2023 Mar;20(3):853-860. [doi: 10.1111/iwj.13932] [Medline: 36054243]
35. Reyzelman AM, Shih CD, Tovmassian G, et al. An evaluation of real-world smart sock-based temperature monitoring data as a physiological indicator of early diabetic foot injury: case-control study. JMIR Form Res 2022 Apr 1;6(4):e31870. [doi: 10.2196/31870] [Medline: 35363148]
36. Scholten HJ, Shih CD, Ma R, Malhotra K, Reyzelman AM. Utilization of a smart sock for the remote monitoring of patients with peripheral neuropathy: cross-sectional study of a real-world registry. JMIR Form Res 2022 Mar 1;6(3):e32934. [doi: 10.2196/32934] [Medline: 35230248]
37. Samarentsis AG, Makris G, Spinthaki S, Christodoulakis G, Tsiknakis M, Pantazis AK. A 3D-printed capacitive smart insole for plantar pressure monitoring. Sensors (Basel) 2022 Dec 12;22(24):9725. [doi: 10.3390/s22249725] [Medline: 36560095]
38. Beach C, Cooper G, Weightman A, Hodson-Tole EF, Reeves ND, Casson AJ. Monitoring of dynamic plantar foot temperatures in diabetes with personalised 3D-printed wearables. Sensors (Basel) 2021 Mar 2;21(5):1717. [doi: 10.3390/s21051717] [Medline: 33801346]

XSL•FO
RenderX

39. Du C, Wang H, Chen H, et al. The feasibility and effectiveness of wearable sensor technology in the management of elderly diabetics with foot ulcer remission: a proof-of-concept pilot study with six cases. Gerontology 2021;67(4):493-502. [doi: 10.1159/000513729] [Medline: 33657570]
40. Abbott CA, Chatwin KE, Foden P, et al. Innovative intelligent insole system reduces diabetic foot ulcer recurrence at plantar sites: a prospective, randomised, proof-of-concept study. Lancet Digit Health 2019 Oct;1(6):e308-e318. [doi: 10.1016/S2589-7500(19)30128-1] [Medline: 33323253]
41. Huchegowda R, Shruti A, Amarendra S, Shraddha T, Huchegowda C. Integrative studies to design and validate wearable footwear among neuropathic patients. Diabetes Metab Syndr 2019 May;13(3):2075-2079. [doi: 10.1016/j.dsx.2019.03.038]
42. Reyzelman AM, Koelewyn K, Murphy M, et al. Continuous temperature-monitoring socks for home use in patients with diabetes: observational study. J Med Internet Res 2018 Dec 17;20(12):e12460. [doi: 10.2196/12460] [Medline: 30559091]
43. Zhou H, Al-Ali F, Rahemi H, et al. Hemodialysis impact on motor function beyond aging and diabetes-objectively assessing gait and balance by wearable technology. Sensors (Basel) 2018 Nov 14;18(11):3939. [doi: 10.3390/s18113939] [Medline: 30441843]
44. Coates J, Chipperfield A, Clough G. Wearable multimodal skin sensing for the diabetic foot. Electronics (Basel) 2016;5(3):45. [doi: 10.3390/electronics5030045]
45. Grewal GS, Schwenk M, Lee-Eng J, et al. Sensor-based interactive balance training with visual joint movement feedback for improving postural stability in diabetics with peripheral neuropathy: a randomized controlled trial. Gerontology 2015;61(6):567-574. [doi: 10.1159/000371846] [Medline: 25721132]
46. Wrobel JS, Ammanath P, Le T, et al. A novel shear reduction insole effect on the thermal response to walking stress, balance, and gait. J Diabetes Sci Technol 2014 Nov;8(6):1151-1156. [doi: 10.1177/1932296814546528] [Medline: 25107709]
47. Banks JL, Petersen BJ, Rothenberg GM, Jong AS, Page JC. Use of a Remote Temperature Monitoring Mat for the Early Identification of Foot Ulcers. Wounds 2020 Feb;32(2):44-49. [Medline: 32155121]
48. Tiwari B, Ntella SL, Jeanmonod K, et al. A polyester–nylon blend plantar pressure sensing insole for person with diabetes. IEEE Sens Lett 2024;8(1):1-4. [doi: 10.1109/LSENS.2023.3338374]
49. Schmalen PG, Meyer T. Temperature sensing insoles for diabetic foot ulcer diagnostics. IEEE Sens Lett 2024;8(7):1-4. [doi: 10.1109/LSENS.2024.3407589]
50. Hemler SL, Ntella SL, Jeanmonod K, et al. Intelligent plantar pressure offloading for the prevention of diabetic foot ulcers and amputations. Front Endocrinol (Lausanne) 2023;14:1166513. [doi: 10.3389/fendo.2023.1166513] [Medline: 37469988]
51. Shirzaei Sani E, Xu C, Wang C, et al. A stretchable wireless wearable bioelectronic system for multiplexed monitoring and combination treatment of infected chronic wounds. Sci Adv 2023 Mar 24;9(12):eadf7388. [doi: 10.1126/sciadv.adf7388] [Medline: 36961905]
52. Thi Ngoc Nga D, Mattana G, Thu VT, Roussel R, Piro B. A simple flexible printed capacitive pressure sensor for chronic wound monitoring. Sens Actuators A Phys 2022 May;338:113490. [doi: 10.1016/j.sna.2022.113490]
53. Brooks E, Burns M, Ma R, Scholten HJ, Becker S. Remote diabetic foot temperature monitoring for early detection of diabetic foot ulcers: a cost-effectiveness analysis. Clinicoecon Outcomes Res 2021;13:873-881. [doi: 10.2147/CEOR.S322424] [Medline: 34675567]
54. De Pascali C, Francioso L, Giampetruzzi L, et al. Modeling, fabrication and integration of wearable smart sensors in a monitoring platform for diabetic patients. Sensors (Basel) 2021 Mar 6;21(5):1847. [doi: 10.3390/s21051847] [Medline: 33800949]
55. Macdonald EM, Perrin BM, Cleeland L, Kingsley MIC. Podiatrist-delivered health coaching to facilitate the use of a smart insole to support foot health monitoring in people with diabetes-related peripheral neuropathy. Sensors (Basel) 2021 Jun 9;21(12):3984. [doi: 10.3390/s21123984] [Medline: 34207743]
56. Cüreklibatır Encan B, Marmaralı A, Bayraktar D, Çetinkalp Ş, Şimşir IY. Clinical performance evaluation of a newly developed sock for people with diabetes. Proc Inst Mech Eng H 2021 Jun;235(6):701-708. [doi: 10.1177/09544119211002074] [Medline: 33730958]
57. Zhang H, Chen Z, Guo S, et al. Sensock: 3D foot reconstruction with flexible sensors. 2020 Apr 21 Presented at: CHI '20 URL: https://dl.acm.org/doi/proceedings/10.1145/3313831 [accessed 2026-06-17] [doi: 10.1145/3313831.3376387]
58. Ngo O, Niemann E, Gunasekaran V, et al. Development of low frequency (20-100 kHz) clinically viable ultrasound applicator for chronic wound treatment. IEEE Trans Ultrason Ferroelectr Freq Control 2019 Mar;66(3):572-580. [doi: 10.1109/TUFFC.2018.2836311] [Medline: 29993739]
59. Lin X, Seet BC. Battery-free smart sock for abnormal relative plantar pressure monitoring. IEEE Trans Biomed Circuits Syst 2017 Apr;11(2):464-473. [doi: 10.1109/TBCAS.2016.2615603] [Medline: 28114035]
60. Lo WT, Yick KL, Ng SP, Yip J. New methods for evaluating physical and thermal comfort properties of orthotic materials used in insoles for patients with diabetes. J Rehabil Res Dev 2014;51(2):311-324. [doi: 10.1682/JRRD.2013.01.0012] [Medline: 24933729]
61. Perrier A, Vuillerme N, Luboz V, et al. Smart diabetic socks: embedded device for diabetic foot prevention. IRBM 2014 Apr;35(2):72-76. [doi: 10.1016/j.irbm.2014.02.004]
62. Garland NT, Song JW, Ma T, et al. A miniaturized, battery-free, wireless wound monitor that predicts wound closure rate early. Adv Healthc Mater 2023 Nov;12(28):e2301280. [doi: 10.1002/adhm.202301280] [Medline: 37407030]

XSL•FO
RenderX

63. Chaudon MJ, Hulea O, Yakoub A, Monnier P, Saadaoui M. Wearable device for iontophoretic treatment and monitoring of pressure ulcers: Proof-of-concept. Med Eng Phys 2022 Sep;107:103861. [doi: 10.1016/j.medengphy.2022.103861] [Medline: 36068041]
64. Das R, Kumar N. Investigations on postural stability and spatiotemporal parameters of human gait using developed wearable smart insole. J Med Eng Technol 2015 Jan;39(1):75-78. [doi: 10.3109/03091902.2014.968676] [Medline: 25350821]
65. Dong T, Guo Y, Gu Y, et al. Design of a wireless and fully flexible insole using a highly sensitive pressure sensor for gait event detection. Meas Sci Technol 2021 Oct 1;32(10):105109. [doi: 10.1088/1361-6501/ac0372]
66. El-Nahas M, El-Shazly S, El-Gamel F, Motawea M, Kyrillos F, Idrees H. Relationship between skin temperature monitoring with smart socks and plantar pressure distribution: a pilot study. J Wound Care 2018 Aug 2;27(8):536-541. [doi: 10.12968/jowc.2018.27.8.536] [Medline: 30086248]
67. Lou D, Pang Q, Pei X, et al. Flexible wound healing system for pro-regeneration, temperature monitoring and infection early warning. Biosens Bioelectron 2020 Aug 15;162:112275. [doi: 10.1016/j.bios.2020.112275] [Medline: 32392156]
68. Pradhan GB, Bhatta T, Shrestha K, Sharma S, Park JY. A wearable, self-sustainable, and wireless plantar pressure and temperature monitoring system for foot ulceration prognosis and rehabilitation. Sens Actuators A Phys 2024 Dec;379:115985. [doi: 10.1016/j.sna.2024.115985]
69. Killeen AL, Brock KM, Dancho JF, Walters JL. Remote temperature monitoring in patients with visual impairment due to diabetes mellitus: a proposed improvement to current standard of care for prevention of diabetic foot ulcers. J Diabetes Sci Technol 2020 Jan;14(1):37-45. [doi: 10.1177/1932296819848769] [Medline: 31122064]
70. Teschner MJ, Seybold BA, Malone BJ, Hüning J, Schreiner CE. Effects of signal-to-noise ratio on auditory cortical frequency processing. J Neurosci 2016 Mar 2;36(9):2743-2756. [doi: 10.1523/JNEUROSCI.2079-15.2016] [Medline: 26937012]
71. Kumbhar S, Bhatia M. Advancements and best practices in diabetic foot care: a comprehensive review of global progress. Diabetes Res Clin Pract 2024 Nov;217:111845. [doi: 10.1016/j.diabres.2024.111845] [Medline: 39243866]
72. Jager W. The importance of clinical trials in advancing medical research. Ann Clin Trials Vaccines Res 2023;13(3):93-98. [doi: 10.37532/ACTVR.2023.13(3).93-98]
73. Raghav A. Diabetic Foot-Recent Advances: IntechOpen; 2023:1-156.
74. Jones P, Bibb R, Davies M, et al. Prediction of diabetic foot ulceration: the value of using microclimate sensor arrays. J Diabetes Sci Technol 2020 Jan;14(1):55-64. [doi: 10.1177/1932296819877194] [Medline: 31596145]
75. Ren Y, Wang H, Song X, Wu Y, Lyu Y, Zeng W. Advancements in diabetic foot insoles: a comprehensive review of design, manufacturing, and performance evaluation. Front Bioeng Biotechnol 2024;12:1394758. [doi: 10.3389/fbioe.2024.1394758] [Medline: 39076210]
76. Billings J, Gee J, Ghulam Z, Abdullah HA. Smart compression sock for early detection of diabetic foot ulcers. Sensors (Basel) 2024 Oct 29;24(21):6928. [doi: 10.3390/s24216928] [Medline: 39517824]
77. Bus SA, Lavery LA, Monteiro-Soares M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update). Diabetes Metab Res Rev 2020 Mar;36 Suppl 1:e3269. [doi: 10.1002/dmrr.3269] [Medline: 32176451]
78. Wang X, Yuan CX, Xu B, Yu Z. Diabetic foot ulcers: classification, risk factors and management. World J Diabetes 2022 Dec 15;13(12):1049-1065. [doi: 10.4239/wjd.v13.i12.1049] [Medline: 36578871]
79. Kaselimi M, Protopapadakis E, Doulamis A, Doulamis N. A review of non-invasive sensors and artificial intelligence models for diabetic foot monitoring. Front Physiol 2022;13:924546. [doi: 10.3389/fphys.2022.924546] [Medline: 36338484]
80. Matijevich E, Minty E, Bray E, Bachus C, Hajizadeh M, Liden B. A multi-faceted digital health solution for monitoring and managing diabetic foot ulcer risk: a case series. Sensors (Basel) 2024 Apr 23;24(9):2675. [doi: 10.3390/s24092675] [Medline: 38732781]
81. Saliba Thorne C, Gatt A, DeRaffaele C, Bazena A, Formosa C. Digital foot health technology and diabetic foot monitoring: a systematic review. Diabetes Res Clin Pract 2021 May;175:108783. [doi: 10.1016/j.diabres.2021.108783] [Medline: 33775686]
82. Gordon IL, Rothenberg GM, Lepow BD, et al. Accuracy of a foot temperature monitoring mat for predicting diabetic foot ulcers in patients with recent wounds or partial foot amputation. Diabetes Res Clin Pract 2020 Mar;161:108074. [doi: 10.1016/j.diabres.2020.108074] [Medline: 32109516]
83. Lockhart M, Dinneen SF, O'Keeffe DT. Plantar pressure measurement in diabetic foot disease: a scoping review. J Diabetes Investig 2024 Aug;15(8):990-999. [doi: 10.1111/jdi.14215] [Medline: 38634342]
84. Castro-Martins P, Marques A, Coelho L, Vaz M, Baptista JS. In-shoe plantar pressure measurement technologies for the diabetic foot: a systematic review. Heliyon 2024 May 15;10(9):e29672. [doi: 10.1016/j.heliyon.2024.e29672] [Medline: 38699042]
85. Lounis K, Zulkernine M. Bluetooth low energy makes "just works" not work. In: Zulkernine M, Zulkernine M, editors. Presented at: 2019 3rd Cyber Security in Networking Conference (CSNet); Oct 23-25, 2019. [doi: 10.1109/CSNet47905.2019.9108931]
86. Omre AH, Keeping S. Bluetooth low energy: wireless connectivity for medical monitoring. J Diabetes Sci Technol 2010 Mar 1;4(2):457-463. [doi: 10.1177/193229681000400227] [Medline: 20307407]

XSL•FO
RenderX

87. Najafi B, Ron E, Enriquez A, Marin I, Razjouyan J, Armstrong DG. Smarter sole survival: will neuropathic patients at high risk for ulceration use a smart insole-based foot protection system? J Diabetes Sci Technol 2017 Jul;11(4):702-713. [doi: 10.1177/1932296816689105] [Medline: 28627227]
88. Bus SA, Aan de Stegge WB, van Baal JG, Busch-Westbroek TE, Nollet F, van Netten JJ. Effectiveness of at-home skin temperature monitoring in reducing the incidence of foot ulcer recurrence in people with diabetes: a multicenter randomized controlled trial (DIATEMP). BMJ Open Diabetes Res Care 2021 Sep;9(1):e002392. [doi: 10.1136/bmjdrc-2021-002392] [Medline: 34493496]
89. Deswal P. NXTSTIM's wearable nerve stimulation device helps manage long-term pain. ClinicalTrials Arena. 2024. URL: https://www.clinicaltrialsarena.com/news/nxtstims-wearable-nerve-stimulation-device-helps-manage-long-term-pain [accessed 2026-06-17]
90. Thomas EE, Taylor ML, Banbury A, et al. Factors influencing the effectiveness of remote patient monitoring interventions: a realist review. BMJ Open 2021 Aug 25;11(8):e051844. [doi: 10.1136/bmjopen-2021-051844] [Medline: 34433611]
91. Ezeamii VC, Okobi OE, Wambai-Sani H, et al. Revolutionizing healthcare: how telemedicine is improving patient outcomes and expanding access to care. Cureus 2024 Jul;16(7):e63881. [doi: 10.7759/cureus.63881] [Medline: 39099901]

Abbreviations

BLE: Bluetooth low energy
DFU: diabetic foot ulcer
MMAT: Mixed Methods Appraisal Tool
NHS: National Health Service
PICO: Population, Intervention, Comparison, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SID: Scientific Information Database

Edited by I Steenstra; submitted 19.Sep.2025; peer-reviewed by A Malik, B Cahill, BA Liden, VFD Carvalho; revised version received 16.May.2026; accepted 19.May.2026; published 13.Jul.2026.

Please cite as:
Ehtemam H, Salehinejad S, Robbins D, Sanaei A, Shirvani H, Sadeghi-Esfahlani S
Wearable Technologies for Diabetic Foot Ulcer Monitoring and Risk Prediction: Systematic Review
JMIR Diabetes 2026;11:e84463
URL: https://diabetes.jmir.org/2026/1/e84463
doi:10.2196/84463

© Houriyeh Ehtemam, Simin Salehinejad, Dan Robbins, Alireza Sanaei, Hassan Shirvani, Shabnam Sadeghi-Esfahlani. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 13.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Based Models for Diabetic Foot Ulcer Assessment: Scoping Review

Muhamad Zulfiqar[1], MSN; Saldy Yusuf[1,2], PhD; Muhammad Jufri Taming[1], MSN; Herlina Burhan[1], BSN

[1]Faculty of Nursing, Hasanuddin University, Jl. Perintis Kemerdekaan KM 10, Makassar, South Sulawesi, Indonesia

[2]Indonesian Diabetic Foot Care Research Group (IndoFootCare), Hasanuddin University, Makassar, South Sulawesi, Indonesia

Corresponding Author:
Saldy Yusuf, PhD
Faculty of Nursing, Hasanuddin University, Jl. Perintis Kemerdekaan KM 10, Makassar, South Sulawesi, Indonesia

Abstract

Background: Diabetic foot ulcers (DFU) are serious complications of diabetes that contribute substantially to morbidity, mortality, and health care burden. Accurate and timely wound assessment is essential for effective DFU management; however, conventional assessment methods are limited by subjectivity, time constraints, and interobserver variability.

Objective: This scoping review aimed to map and synthesize evidence regarding the development and application of artificial intelligence (AI)–based models for DFU assessment.

Methods: A scoping review was conducted following the Arksey and O'Malley framework and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Literature searches were performed in PubMed, ProQuest, and Scopus for studies published between 2014 and 2026. Study selection and data charting were conducted independently by two reviewers using predefined inclusion criteria based on the PCC (population, concept, context) framework. Extracted data were synthesized narratively and categorized according to major AI application domains.

Results: A total of 654 records were identified, of which 46 studies met the inclusion criteria. The included studies predominantly focused on image segmentation, diagnostic classification, and risk prediction or monitoring of DFUs. Convolutional neural networks were the most commonly applied models, with performance evaluated using metrics such as accuracy, Dice similarity coefficient, and area under the curve. Most studies relied on retrospective, single-center datasets, with limited external validation and minimal real-world clinical implementation.

Conclusions: AI-based models demonstrate strong potential to enhance DFU assessment and monitoring by improving accuracy and efficiency. However, significant gaps remain in terms of dataset diversity, external validation, and integration into clinical workflows. Future research should prioritize prospective validation, standardized datasets, and real-world implementation to support safe and effective clinical adoption.

(JMIR Diabetes 2026;11:e77925) doi:10.2196/77925

KEYWORDS

artificial intelligence; deep learning; diabetic foot ulcers; diabetic wound; scoping review

Introduction

Diabetic foot ulcers (DFUs) significantly increase the risk of morbidity and mortality in patients with diabetes. The lifetime incidence of DFU ranges from 19% to 34%, with an annual incidence of 2%. Although DFU can heal, the recurrence rate remains high: at 40% within 1 year and 65% within 3 years [1]. DFUs are associated with significant morbidity and mortality, with a 5-year mortality risk that is 2.5 times higher than that of diabetic patients without DFU [2]. The management of DFU faces significant challenges, including the complexity of their causes, slow healing processes, and high risk of complications, such as infections and amputations [3]. Additionally, patients with DFUs often experience a considerable decline in their quality of life, including chronic pain, mobility limitations, and psychological distress, such as anxiety and depression [4]. Therefore, appropriate DFU management, particularly wound assessment, is crucial for successful treatment.

Wound assessment is the initial step in managing DFU. Accurate and comprehensive wound assessment is critical for wound management [5]. Wound assessment tools provide scores or values that reflect changes in the clinical condition [6]. Several wound assessment tools are available, including the Bates-Jensen Wound Assessment Tools [7], the Diabetic Foot Ulcer Assessment Scale (DFUAS) [8], the DMIST (depth, maceration, inflammation/infection, size, tissue type of the wound bed, type of wound edge, and tunneling/undermining) tool [9], and the DEPA (depth, extent of bacterial colonization,

XSL•FO
RenderX

phase of healing, and associated aetiology) tool [10]. Typically, wound assessment and monitoring are performed by nurses with specialized training in wound care [11]. Wound assessment depends on health care providers' clinical experience, which may lead to variations in diagnosis and treatment [12]. Some studies have shown that insufficient documentation in wound assessment can hinder the early detection of complications and negatively affect treatment outcomes [13]. Furthermore, conventional methods of assessing the severity of DFU are often time-consuming, labor-intensive, and prone to discrepancies, which hampers accurate patient monitoring [14]. Another challenge is the technological limitations in wound assessment, as manual wound documentation is often inaccurate and can lead to inconsistencies in care [15]. Therefore, innovation is needed to enhance the efficiency and objectivity of wound assessment.

The field of artificial intelligence (AI) is rapidly expanding, particularly in health care. AI can be applied to diagnose diseases, design personalized treatment plans, and assist clinicians in decision-making [16]. With technological advancements, several commercially available wound assessment or monitoring systems are now available to track chronic wounds [17]. AI can help improve wound assessment accuracy, patient engagement, and compliance with wound care regimens [18]. AI has the potential to significantly enhance the early detection and management of DFU.

Accurate and objective assessment of DFU remains challenging in clinical practice. Recently, AI approaches using digital image analysis have been increasingly explored to support DFU screening and evaluation. Previous systematic reviews have reported that convolutional neural network (CNN)–based models are the most commonly used approaches for DFU image segmentation and screening [19]. Furthermore, AI has proven to be more accurate than conventional methods for wound image analysis [20]. The advancement of AI technology allows for the development of more accurate and efficient wound assessment models with deep learning algorithms that achieve accuracy comparable to that of nurses in diagnosing DFU. These systems provide reliable assessments to formulate personalized care plans, aligning with research showing machine learning algorithms' potential in DFU recognition [21]. AI has demonstrated promising results in the classification and localization of DFU, with high accuracy in identifying ischemia and infection, ranging from 73% to 95.4% [22]. Wound assessment systems, ranging from computer-based to mobile applications, have the potential to integrate AI to enhance measurement accuracy [23]. Despite several commercially available wound assessment systems, most of these systems have not been reviewed in the literature regarding measurement accuracy, mainly since DFU may occur in curved or angled areas on the foot [17]. Therefore, this scoping review aimed to map and synthesize evidence regarding the development and application of AI-based models for DFU assessment.

Methods

Scoping Review Framework

This scoping review was conducted in accordance with the 5-stage framework proposed by Arksey and O'Malley [24] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [25]. This review aimed to map and synthesize evidence regarding the development and application of AI-based models for DFU assessment.

The scoping review methodology enables systematic mapping of the existing literature, identification of research gaps, and synthesis of available evidence. The methodological framework consisted of five stages: (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) data charting, and (5) collating, summarizing, and reporting the results.

Stage I: Research Question

The research question guiding this review was: How have AI-based models been developed and applied for diabetic foot ulcer assessment, and what are their methodological characteristics and reported-performance outcomes?

Stage II: Identifying Relevant Studies

A comprehensive literature search was conducted in PubMed, ProQuest, and Scopus to identify relevant studies published between 2014 and 2026. The search strategy included combinations of keywords related to diabetic foot ulcers, artificial intelligence, machine learning, deep learning, and wound assessment. The detailed search strings for each database are provided in Multimedia Appendix 1. The eligibility criteria were defined using the PCC (population, concept, context) framework recommended by the Joanna Briggs Institute.

Population included studies involving individuals with DFU, without restriction on age, sex, or disease severity. *Concept* focused on studies describing the development or application of AI-based models, including machine learning, deep learning, and computer vision approaches, for DFU assessment, encompassing tasks such as segmentation, classification, risk prediction, monitoring, and decision support. *Context* included studies conducted in clinical, experimental, or health care–related technological settings, without geographical restriction.

Only peer-reviewed full-text articles published in English between 2014 and 2026 were included. Efforts were made to obtain full-text articles through institutional access and other available sources to minimize the risk of exclusion based on access limitations.

Stage III: Study Selection

All retrieved records were imported into Rayyan for management and screening purposes. Duplicate references were identified and removed prior to review [26]. Title and abstract screenings were conducted independently by two reviewers. Studies deemed potentially eligible proceeded to independent full-text assessment by the same reviewer. Disagreements at

XSL•FO
RenderX

any stage of the screening process were resolved through discussion for agreement. When agreement was not reached, a third reviewer was consulted.

A total of 654 records were identified through database searches, including PubMed (n=245), ProQuest (n=123), and Scopus (n=286). After removing 310 duplicate records, 344 records remained and were screened based on titles and abstracts, of which 196 were excluded. Subsequently, 148 reports were sought for retrieval. Of these, 41 reports could not be retrieved after attempts to access full texts through available institutional and alternative sources. As a result, 107 full-text articles were assessed for eligibility. Following full-text assessment, 61 articles were excluded because they were not related to DFU. Ultimately, a total of 46 studies were included in this review.

Stage IV: Data Mapping

Data were extracted to capture study characteristics, including author and year, country, study design, dataset characteristics, AI methods, application domain, and reported performance metrics. Information regarding the application of AI-based assessment models for DFU was also extracted.

Stage V: Thematic Summary and Key Findings

The extracted data were synthesized narratively using a deductive thematic approach. This approach was guided by predefined analytical domains developed based on the review objective and commonly reported categories of AI applications in DFU assessment. The predefined domains included segmentation and measurement, diagnostic classification, risk prediction and monitoring, and clinical decision support systems. Two reviewers independently reviewed the extracted data and assigned each study to one or more of the predefined domains based on its primary application focus. During this process, studies were examined for their methodological characteristics, application objectives, and reported outcomes to ensure accurate categorization. Any discrepancies in domain classification were resolved through discussion to reach consensus. When necessary, a third reviewer was consulted to adjudicate disagreements. This process ensured consistency and reliability in the thematic grouping of the included studies. Consistent with the scoping review methodology, no formal risk-of-bias assessment was conducted.

Ethical Considerations

This scoping review was not prospectively registered in any database. As this study synthesized data from publicly available literature and did not involve human participants or identifiable data, formal ethical approval was not required.

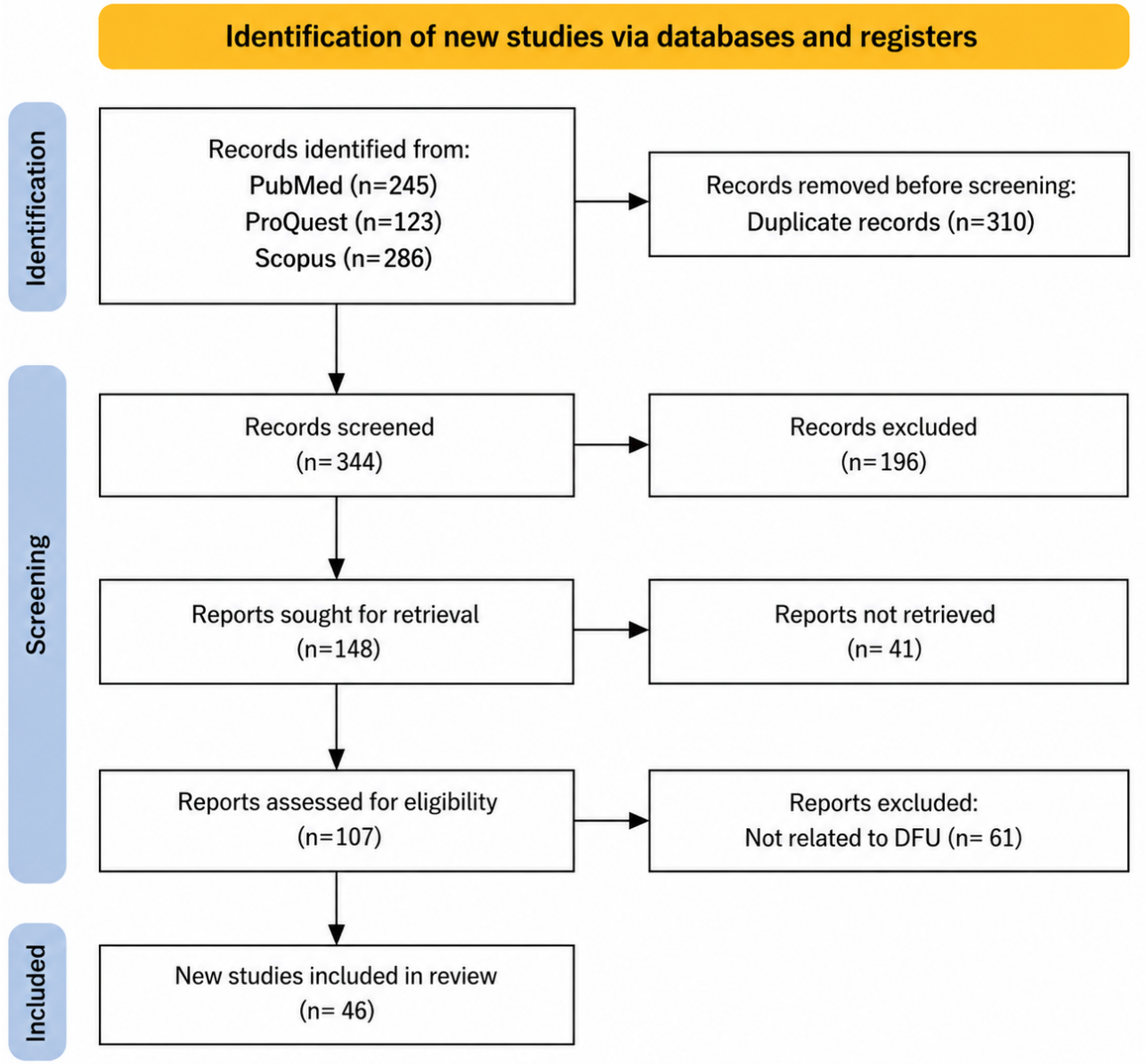
Results

Study Characteristics

A total of 46 studies were included in this review (Figure 1). The key characteristics of the included studies, including wound assessment task, AI methods, data sources, and primary outcomes (Table 1).

XSL•FO
RenderX

Figure 1. Flow diagram of study search and selection based on PRISMA-ScR. DFU: diabetic foot ulcer.



XSL•FO
RenderX

Table . Overview of artificial intelligence (AI) methods and application domains in the included studies.

Characteristics	Studies	AI methods[a]	Data source[b]	Primary outcome
Wound segmentation, measurement, and characterization	[17,27-38]	Fast R-CNN; bi-CNN; DL-based mobile app; SwishRes-U-Net; wound viewer AI; AHRF; AI measurement app; CNN; explainable DL; CNN-based detection; thermal DL; benchmark DL; DL segmentation	DFU image datasets; DFU photos; DFU/FUSEG datasets; clinical patients; multidevice images; public datasets; multi-dataset DFU	Wound localization; depth and granulation measurement; automated wound measurement; tissue boundary extraction; accurate 3D measurement and wound bed preparation classification; edge detection; measurement consistency; accurate wound segmentation and classification; interpretable segmentation; granulation detection; improved detection accuracy; standardized evaluation; generalizability analysis
Diagnostic classification and condition assessment	[14,39-52]	Hybrid DL classifier; FusionNet (XAI); ScoreDFUNet; DenseVAE-CL; ConMatFormer; few-shot DL; multimodal DL; attention DL; hybrid CNN; CNN+Transformer; efficient CNN; lightweight CNN; explainable DL; Transformer+XAI; CNN+ELM	DFU images; 2673 Kaggle images; DFUC2021; Kaggle datasets; clinical images	Condition classification; explainable diagnosis; image-based diagnosis; very high classification accuracy; improved accuracy and explainability; accurate classification in limited datasets; improved classification performance; enhanced feature extraction; efficient and interpretable classification
Risk prediction, complication detection, and longitudinal monitoring	[22,53-65]	LightGBM; AutoML DL; DFUCare platform; DFUHelper; CNN+SPCD; Siamese neural network; Mask R-CNN thermal fusion; ML model; temporal ML; AI monitoring system; deep learning	Clinical datasets; DFU datasets; clinical images; serial DFU images; prospective clinical patients; population datasets; longitudinal data	Outcome prediction; infection and ischemia detection; healing trend monitoring; status differentiation; area measurement and healing monitoring; amputation prediction; healing prediction; diabetic foot ulcer occurrence prediction; healing trajectory prediction; severity monitoring; early complication detection
Clinical decision support and automated model development	[44,66-68]	AI-based DSS; AutoML DL; ML decision model; explainable DL	DFU clinical cases; DFU datasets; clinical datasets	Decision agreement; automated model generation; risk stratification; transparent clinical decision-making

[a]AHRF: adaptive hybrid regression framework; AI: artificial intelligence; AutoML: automated machine learning; bi-CNN: bilinear convolutional neural network; CNN: convolutional neural network; DL: deep learning; DSS: decision support system; ELM: extreme learning machine; LightGBM: light gradient boosting machine; ML: machine learning; R-CNN: region-based convolutional neural network; SPCD: superpixel color descriptor; XAI: explainable artificial intelligence.

[b]DFU: diabetic foot ulcer; DFUC: Diabetic Foot Ulcer Challenge; FUSEG: Foot Ulcer Segmentation Dataset.

Diagnostic classification and condition assessment (n=15, 32.6%) and risk prediction, complication detection, and longitudinal monitoring (n=14, 30.4%) were the most frequently investigated domains, followed by segmentation, measurement, and characterization (n=13, 28.3%). Clinical decision support and automated model development were less frequently studied (n=4, 8.7%) (Table 2). This distribution reflects a strong research emphasis on image-based analysis and predictive modeling, with comparatively limited exploration of clinically integrated decision-support systems.

To further examine the distribution of evidence and identify potential research gaps, an evidence gap map was developed based on AI application domains and key methodological characteristics (Table 3).

XSL•FO
RenderX

Table . Distribution of artificial intelligence application domains in included studies (N=46).

Artificial intelligence application domain	Studies, n (%)
Diagnostic classification and condition assessment	15 (32.6)
Risk prediction, complication detection, and longitudinal monitoring	14 (30.4)
Segmentation, measurement, and characterization	13 (28.3)
Clinical decision support and automated model development	4 (8.7)

Table . Evidence gap map of artificial intelligence (AI) application domains in diabetic foot ulcer assessment.

Characteristics	Segmentation, measurement, and characterization	Diagnostic classification and condition assessment	Risk prediction, complication detection, and longitudinal monitoring	Clinical decision support and automated model development
Image-based diagnosis	✓	✓	✓	✓
Clinical validation datasets	X	✓	✓	✓
Prospective studies	X	X	X	X
External validation	X	X	✓	X
Real-world implementation	X	X	✓	X
Explainable AI (XAI)	✓	✓	✓	✓
Multimodal data	✓	✓	X	X

The evidence gap map illustrates the distribution of studies across AI application domains and key methodological characteristics. The findings demonstrate a strong concentration of evidence in segmentation and diagnostic classification domains, particularly those using image-based datasets, indicating a dominant reliance on computer vision approaches for DFU assessment. In contrast, the use of clinical or tabular data remains limited and is primarily observed in studies focusing on risk prediction and monitoring. Notably, prospective study designs were absent across all domains, and external validation was rarely reported, with only limited evidence identified in prediction-related studies. Furthermore, real-world implementation of AI models remains scarce, with only a small number of monitoring studies demonstrating initial clinical application. Although explainable AI approaches have been increasingly incorporated across domains, their integration remains inconsistent. Overall, these findings highlight that while the technical development of AI models for DFU assessment is rapidly advancing, substantial gaps persist in clinical validation, generalizability, and real-world implementation.

This scoping review mapped the application of AI-based assessment models in DFU. Various studies have been conducted to enhance the detection, evaluation, and monitoring of DFU using AI, machine learning, and computer vision approaches. Some studies have focused on predicting amputation levels in DFU patients by using explainable machine learning models that classify DFU based on the Wagner and WIfI systems while also evaluating risk factors such as the duration of diabetes, history of amputations, and vascular conditions [53]. Additionally, machine learning–based models have been used to identify risk factors for DFU through local foot examination [54]. In wound detection and segmentation, several studies have developed deep learning–based models, such as Fast R-CNN, which enable automatic measurement of wound size, boundaries, and geometry of DFU [27]. A bilinear CNN-based model was also developed for fine-grained wound classification based on depth and granulation tissue [28]. Meanwhile, another study explored explainable AI approaches using algorithms such as SHAP (Shapley Additive Explanations), LIME (local interpretable model-agnostic explanations), and Grad-CAM (gradient-weighted class activation mapping), which enhance transparency in DFU detection and classification [39]. New datasets have also been developed to identify ischemia and infection in DFU using computer vision techniques [55].

Smart application–based technology has been integrated into DFU monitoring through various innovative applications, enabling real-time wound dimension measurement, validation with manual methods, and automatic tissue classification [29]. Furthermore, hardware acceleration approaches such as field-programmable gate array (FPGA) and graphics processing unit (GPU) have been developed to enhance the efficiency of real-time wound classification [40].

Regarding wound boundary determination, the associative hierarchical random field (AHRF) method was introduced as a more accurate alternative to traditional methods by reducing dependence on lighting and camera angles [30]. Another study developed an AI-based automated scoring system to assess DFU severity by considering wound classification, size, and color characteristics [14]. AI-based approaches have also been used to assess amputation levels in DFU patients based on clinical photographs, compared to physician decisions using Wagner classification [66]. Deep learning–based platform development has also been conducted for noninvasive DFU detection and monitoring [22]. The SwishRes-U-Net model has been applied to improve DFU evaluation accuracy [31]. Meanwhile, a Siamese neural network–based framework has been developed for longitudinal DFU evaluation to monitor wound progression under various clinical conditions [56].

XSL•FO
RenderX

Several studies have focused on quantitative wound assessment using advanced imaging technologies. Similarly, an AI-powered medical device (Wound Viewer) has been validated for remote wound assessment, showing strong agreement with clinician evaluations in measuring wound parameters and classifying tissue characteristics [32].

Other studies have primarily focused on image segmentation and severity classification. An AI-enhanced imaging framework combining active contour modeling and deep learning classification has been developed to segment ulcer boundaries and predict severity based on Wagner grading, achieving high segmentation accuracy [57]. In addition, a Mask R-CNN model integrated with thermal and visual image fusion has been applied to evaluate DFU healing trajectories, demonstrating a strong correlation with clinician-based measurements [58].

Several deep learning architectures have been proposed for diagnostic classification. A hybrid deep learning architecture integrating convolutional, attention, and transformer modules has been introduced for DFU classification, reporting high accuracy and improved interpretability through explainable AI methods [67]. Another framework combining DenseNet, variational autoencoders, and contrastive learning was developed to improve the classification performance and model generalization [41]. Additionally, an ensemble deep learning system combining YOLOv8 and Faster R-CNN was developed to enhance DFU detection and localization accuracy, demonstrating improved performance compared with individual models [42].

Automated DFU detection and classification using traditional machine learning approaches has been demonstrated as feasible in early image-based studies [69], providing a foundation for subsequent methodological advancements. Building on this, deep learning–based architectures have further improved classification performance by enabling more robust feature extraction from DFU images [43]. More recently, transformer-based models have enhanced classification accuracy while improving feature representation and generalization across datasets [44], and hybrid architectures combining CNNs and vision transformers have further strengthened DFU detection performance by integrating local and global feature learning [45]. In data-limited scenarios, few-shot learning approaches have enabled effective DFU classification despite constrained labeled datasets [46], while attention-based deep learning models have improved feature extraction and classification accuracy by focusing on clinically relevant regions of the image [47].

Beyond performance improvements, explainable AI methods have enhanced transparency in DFU detection by providing visual explanations of model predictions [48], thereby supporting clinical interpretability. At the same time, lightweight deep learning architectures have demonstrated high performance in DFU detection and grading tasks with reduced computational burden, facilitating real-time applications [49]. Multimodal deep learning approaches integrating RGB (red, green, and blue) and thermal data have further improved classification accuracy by incorporating complementary physiological information [50], while texture-based feature extraction combined with deep learning has also been shown to enhance classification performance [51].

In parallel with classification advancements, deep learning–based segmentation approaches have been widely used to delineate wound boundaries and support quantitative wound assessment [33]. However, cross-dataset evaluation studies have highlighted ongoing challenges in model generalizability, particularly when applied to heterogeneous clinical data [34]. To address this, thermography-based segmentation models have enabled improved wound localization by capturing physiological changes associated with DFU [35], while advanced segmentation models have demonstrated strong performance in identifying multiple tissue types within wounds [36]. Additionally, granulation tissue detection models have been developed to support monitoring of wound healing progression using AI-based approaches [37]. Extending beyond static assessment, temporal machine learning frameworks have enabled longitudinal monitoring of wound healing trajectories across time [59], supporting more dynamic clinical evaluation. Furthermore, predictive models have been developed to estimate the risk of minor amputation in DFU patients using clinical variables [60], and machine learning approaches have also been applied to predict hard-to-heal ulcers and support prognosis assessment [61].

Overall, the included studies demonstrate a rapid expansion of AI applications in DFU assessment, encompassing wound segmentation, diagnostic classification, risk prediction, and longitudinal monitoring. The findings indicate that deep learning and machine learning approaches have significantly improved the accuracy and efficiency of DFU detection and evaluation, particularly through advanced imaging analysis and automated classification systems. However, several challenges remain, including variability in model generalizability across datasets, limited integration into clinical decision-making workflows, and the need for standardized validation using diverse and real-world data. These gaps highlight the necessity for future research to focus not only on model performance but also on clinical applicability, robustness, and implementation to ensure that AI-based DFU assessment tools can be effectively translated into routine clinical practice.

Study Design

The study designs presented in these articles demonstrate various approaches and methodologies used to investigate DFU. Various research methods have been used in studies focusing on the detection, analysis, and management of DFU. One study used machine learning models such as random forest and support vector machine with Monte Carlo cross-validation to develop a DFU risk prediction model [54]. Deep learning methods were also applied in DFU image segmentation using Fast R-CNN and transfer learning to enhance wound detection accuracy [27]. Another study used a bilinear CNN to automatically measure wound depth and granulation tissue as a classification method based on medical images [28]. To evaluate the accuracy of DFU wound area measurements, a comparison was made between innovative applications, ruler methods, and ImageJ software to identify more efficient and accurate solutions [29]. The AHRF

XSL•FO
RenderX

method was also developed to improve the accuracy of wound boundary segmentation under varying lighting conditions [30].

In a study evaluating AI recommendations for determining DFU amputation levels, an analysis was conducted on 60 patients, comparing AI recommendations with medical team decisions [66]. Another study used a SwishRes-U-Net deep learning model, consisting of a dual U-Net and a pretrained SwishResNet, for chronic wound segmentation, including DF [31]. To improve transparency in medical decision-making, the light gradient boosting machine (LightGBM) model, combined with SHAP, was used to develop a machine learning model capable of explaining DFU patient amputation predictions [53]. In real-time DFU wound classification studies, 2 CNN models, DFU_FNet and DFU_TFNet, were applied using hardware acceleration such as FPGA and GPU to enhance diagnostic efficiency [40]. In developing deep learning–based DFU detection platforms, the YOLOv5s model was used for wound segmentation, while other deep learning models were applied for infection and ischemia classification [22]. Moreover, explainable AI methods based on multi-CNN, combining DenseNet201, VGG19, and NASNetMobile, were applied to improve transparency in DFU diagnosis using algorithms like SHAP, LIME, and Grad-CAM [39].

The deep learning method ScoreDFUNet was developed for wound classification based on ulcers, infection, normal skin, and gangrene categories, aiming to enhance consistency in wound assessment [14]. For long-term DFU monitoring, the Siamese neural network method was applied to compare wound conditions over time to understand patient healing progression [56]. Additionally, CNN with superpixel color descriptor (SPCD) was used in research focused on identifying ischemia and infection in medical images to support early DFU diagnosis [55]. In clinical validation of AI-based applications for wound measurement and monitoring, studies were conducted comparing AI measurement results with manual methods performed by wound care nurses to assess the accuracy and reliability of the applications [17]. Finally, deep learning models were again developed using CNN approaches with hardware acceleration to enable more efficient real-time DFU wound classification [40].

The additional studies included in this review demonstrate diverse methodological designs, reflecting the evolving maturity of AI research in DFU assessment. Several studies used retrospective image datasets to train and evaluate deep learning models. A retrospective cohort study using a large dataset of DFU images was conducted to develop and validate segmentation and classification algorithms [57]. Similarly, an experimental model development study used publicly available DFUC (Diabetic Foot Ulcer Challenge) datasets to evaluate a hybrid deep learning classification model [67]. Another study used benchmark DFUC2020 and IEEE datasets to develop and evaluate an ensemble detection model for DFU localization [42]. In addition, a deep learning framework trained on publicly available Kaggle datasets was developed to assess classification performance and model generalization [41].

Prospective and clinical validation designs were also represented. A prospective clinical study evaluated healing trajectories using thermal and visual imaging to assess wound progression [58]. In addition, a comparative clinical validation study involving patients with various chronic wounds evaluated the reliability of an AI-enabled medical device for remote wound assessment [32].

Overall, these methodological approaches include experimental model development studies, retrospective dataset analyses, and prospective clinical validation studies, indicating a transition from purely technical model development toward clinically oriented evaluation of AI systems in DFU care.

Key Findings

Building on the study characteristics described above, Table 4 summarizes each included study in terms of AI model type, dataset characteristics, application focus, and key outcomes.

XSL•FO
RenderX

Table . Synthesis analysis of 23 articles relevant to the objective.

	Studies	AI methods[a]	Datasets[b]	Key findings and summary
Wound segmentation, measurement, and characterization	[17,27-38]	Fast R-CNN; bi-CNN; DL mobile app; SwishRes-U-Net; wound viewer AI; AHRF; AI measurement app; CNN; explainable DL; CNN detection; thermal DL; benchmark DL; DL segmentation	DFU image datasets; DFU photos; multidevice images; in vitro and clinical wound patients; retrospective DFU cohort; DFUC2020 and IEEE datasets; clinical patients; DFU images; thermal DFU images; public datasets; multi-dataset DFU	AI demonstrates strong capability in automated wound detection, segmentation, and measurement, including depth and tissue characterization, with high accuracy and reliability comparable to clinical assessment.
Diagnostic classification and condition assessment	[14,39-52]	Hybrid DL classifier; XAI-FusionNet; ScoreDFUNet; ConMatFormer; Dense-VAE-CL; few-shot DL; multimodal DL; attention DL; hybrid CNN; CNN+Transformer; efficient CNN; lightweight CNN; explainable DL; Transformer+XAI; CNN+ELM	DFU image datasets; DFUC2021 and Kaggle datasets; Kaggle DFU images; DFU images	Deep learning models achieve high accuracy in diagnostic classification and severity assessment, with explainable AI improving interpretability and clinical relevance.
Risk prediction, complication detection, and longitudinal monitoring	[22,53-65]	LightGBM; AutoML DL; DFUCare; DFU-Helper; CNN+SPCD; Siamese NN; Mask R-CNN fusion; ML; temporal ML; AI system; deep learning	Clinical datasets; DFU datasets; clinical images; serial DFU images; prospective clinical patients; clinical dataset; population dataset; longitudinal data; clinical images; DFU images	AI models effectively predict clinical outcomes, detect complications such as infection and ischemia, and enable longitudinal monitoring of wound healing with strong correlation to clinician assessment.
Clinical decision support and automated model development	[44,66-68]	AI-based DSS; AutoML DL; ML DSS; explainable DL	DFU clinical cases; DFU datasets; clinical dataset; DFU images	AI supports clinical decision-making with good agreement with physicians and enables automated predictive model development with high performance.

[a]AHRF: adaptive hybrid regression framework; AI: artificial intelligence; AutoML: automated machine learning; bi-CNN: bilinear convolutional neural network; CNN: convolutional neural network; DL: deep learning; DSS: decision support system; ELM: extreme learning machine; LightGBM: light gradient boosting machine; Mask R-CNN: mask region-based convolutional neural network; ML: machine learning; NN: neural network; R-CNN: region-based convolutional neural network; SPCD: superpixel color descriptor; XAI: explainable artificial intelligence.

[b]DFU: diabetic foot ulcer; DFUC: Diabetic Foot Ulcer Challenge.

Wound Segmentation, Measurement, and Characterization

A substantial body of evidence focused on the application of AI for wound segmentation, measurement, and characterization. Deep learning models such as Fast R-CNN demonstrated strong capability in detecting wound boundaries from medical images [27], while bilinear CNNs improved estimation of wound depth and granulation tissue compared with conventional approaches [28]. Probabilistic frameworks such as the AHRF enhanced segmentation accuracy under variable imaging conditions [30], and architectures such as SwishRes-U-Net further demonstrated robust performance across multiple datasets [31].

AI-powered medical devices showed strong reliability for remote wound assessment and tissue classification [32]. Validation studies of smart applications indicated that AI-based tools can provide efficient wound measurements with comparable accuracy to manual methods [29]. Additional evaluations demonstrated that AI-enabled measurement devices may reliably replace traditional measurement approaches in a clinical setting [17].

Deep learning–based diagnostic models achieved high accuracy in simultaneous wound segmentation and classification, thereby improving overall assessment efficiency [36]. Building on this, explainability-integrated models further enhanced segmentation performance while maintaining interpretable identification of wound regions, supporting clinical transparency [38]. In addition to structural segmentation, automated detection models demonstrated strong performance in identifying granulation tissue, enabling more objective evaluation of wound healing progression [37]. Furthermore, thermography-based segmentation frameworks improved detection accuracy by capturing temperature-related physiological changes associated with tissue damage [35]. Consistent with these advancements, benchmark evaluations demonstrated that segmentation models achieved competitive and reliable performance across standardized DFU datasets [33]. However, cross-dataset validation studies revealed reduced performance when applied to external datasets, highlighting ongoing limitations in model

XSL•FO
RenderX

generalizability [34]. Overall, these findings indicate that AI-based segmentation and measurement technologies can improve objectivity and consistency in wound assessment.

Diagnostic Classification and Condition Assessment

Another major domain involved diagnostic classification and condition assessment. Ensemble frameworks combining multiple detection models demonstrated improved localization and diagnostic performance with high precision [42], while generative contrastive frameworks integrating DenseNet and variational autoencoders improved model generalization and classification accuracy across datasets [41].

Explainable AI approaches such as FusionNet enabled automated detection and analysis of DFU while providing interpretable outputs to support clinical interpretation [39]. The ScoreDFUNet model demonstrated strong capability in classifying DFU images into clinically relevant categories [14]. Real-time classification supported by hardware acceleration platforms further highlighted the feasibility of AI deployment in clinical environments [40].

Few-shot learning models achieved robust classification performance despite limited training data, demonstrating strong efficiency in data-scarce conditions [46]. Building on this, multimodal deep learning frameworks further improved classification accuracy by integrating complementary feature representations from multiple data sources [50]. In parallel, attention-based architectures achieved superior feature extraction, leading to improved classification accuracy compared with conventional models [47]. Similarly, hybrid feature extraction approaches combining texture descriptors with deep learning features enhanced classification performance [51]. Extending this approach, hybrid CNN–vision transformer models demonstrated superior performance by effectively capturing both local and global image features [45]. Moreover, efficient deep learning architectures achieved high classification accuracy while reducing computational complexity, supporting more practical implementation [43]. In line with this, lightweight architectures enabled accurate real-time DFU detection and grading with minimal computational burden [49]. Beyond performance, explainable AI frameworks improved diagnostic transparency by providing interpretable visual explanations without compromising accuracy [48]. Complementing these findings, transformer-based explainable models further enhanced interpretability while maintaining high classification performance suitable for clinical use [44]. Finally, hybrid CNN-ELM approaches achieved competitive classification performance while reducing training complexity, offering an efficient alternative for resource-limited settings [52]. Collectively, these findings indicate increasing maturity of AI-based diagnostic classification systems for DFU assessment.

Risk Prediction, Complication Detection, and Longitudinal Monitoring

Several studies focused on risk prediction, complication detection, and longitudinal monitoring of DFU. Machine learning models demonstrated the ability to estimate clinical risk such as amputation likelihood and disease progression, with moderate-to-high discriminatory performance across metrics including accuracy, sensitivity, specificity, and area under the curve [54]. However, evidence of external validation remained limited, with most studies relying on internal testing datasets [53]. Severity classification frameworks integrating imaging and predictive modeling also demonstrated high accuracy in predicting ulcer severity and supporting clinical risk stratification [57].

AI-based monitoring approaches further supported longitudinal assessment of wound progression. The DFUCare model enabled automated detection and monitoring of wound healing and complications such as infection and ischemia [22], while CNN approaches incorporating SPCDs demonstrated the capability in detecting ischemia and infection [55]. Siamese neural network frameworks enabled comparison of wound conditions over time to track healing trajectories [56], and image fusion models integrating thermal and visual imaging showed strong agreement with clinician measurements in monitoring healing progression [58].

Machine learning models achieved high predictive accuracy in estimating minor amputation risk, thereby supporting more informed clinical decision-making [60]. Building on this, predictive models demonstrated strong discriminative ability in identifying hard-to-heal DFUs at early stages, enabling earlier intervention strategies [61]. Furthermore, externally validated models achieved reliable performance in predicting DFU infection across different clinical settings, indicating good generalizability [62]. At the population level, predictive models also achieved strong accuracy in identifying individuals at risk of developing DFU, supporting preventive care approaches [63].

In addition to risk prediction, temporal machine learning frameworks achieved high accuracy in predicting wound healing trajectories across multiple clinical visits, allowing for dynamic monitoring of patient outcomes [59]. Consistent with this, AI-based monitoring systems demonstrated strong agreement with clinical assessments in evaluating wound severity and progression, reinforcing their clinical applicability [64]. Moreover, thermography-based deep learning models improved early detection sensitivity for DFU-related complications by capturing physiological changes, further enhancing early diagnostic capability [65]. These findings highlight the potential role of AI in supporting early risk identification and objective monitoring of DFU progression.

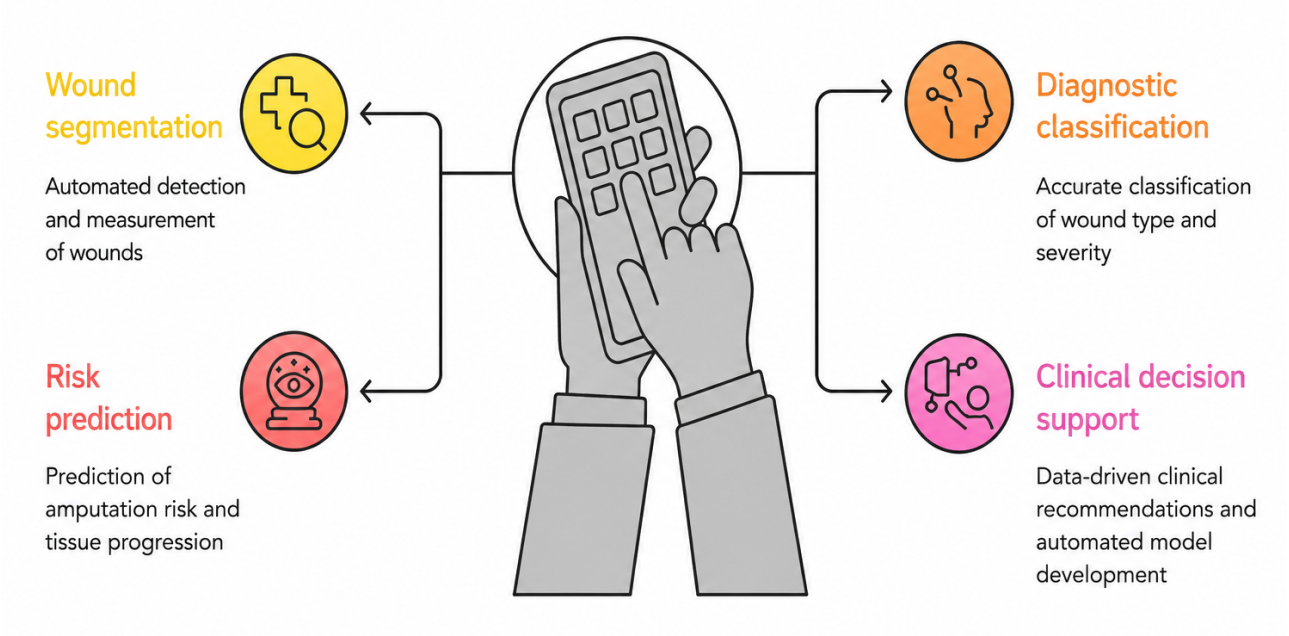
Clinical Decision Support and Automated Model Development

AI has also been applied to support clinical decision-making and automated model development. Studies reported that AI-generated recommendations for amputation level were generally consistent with physician judgment, indicating potential for decision support in clinical practice [66]. Explainability and transparency remain important components of decision support systems. Hybrid deep learning frameworks incorporating explainable AI methods improved interpretability of predictions and supported clinician understanding of decision pathways [67]. Machine learning–based clinical decision models achieved high predictive performance in supporting risk stratification and treatment planning, thereby enhancing the precision of clinical decision-making in DFU management [68].

XSL•FO
RenderX

Building on this, explainability-driven deep learning approaches further improved model transparency while maintaining strong diagnostic performance, thereby increasing clinician trust and facilitating the integration of AI-based decision systems into clinical practice [44]. Overall, these findings suggest that AI-based decision support tools may enhance clinical workflow efficiency and support evidence-based DFU management. The specific AI approaches and their clinical application domains were mapped to provide an overview of how AI has been used to support DFU assessment (Figure 2).

Figure 2. Key benefits of artificial intelligence in diabetic foot ulcer assessment. AI: artificial intelligence.



Discussion

Principal Findings

Wound Segmentation, Measurement, and Characterization

Deep learning models have increasingly been applied to the detection and segmentation of DFU, offering a more standardized and reproducible alternative to conventional manual assessment. These models enable automated boundary delineation, depth classification, and tissue identification, which are key elements in DFU management [29,30].

The growing use of real-time and hardware-accelerated models such as DFU_TFNet further indicates the potential for deployment in point-of-care environments [54]. Performance is increasingly evaluated using standardized segmentation metrics such as Dice score and intersection over union (IoU), allowing comparison across architectures including SwishRes-U-Net, SegNet, and U-Net [55].

Recent studies also demonstrate continued progress in AI-enabled wound measurement technologies. An imaging-based AI system integrating advanced sensing technologies such as LiDAR (light detection and ranging) demonstrated high accuracy in quantifying wound dimensions and tissue composition, supporting its potential role as an objective assessment tool [57]. An AI-powered wound imaging device validated in clinical settings demonstrated strong agreement with clinician measurements, reinforcing its reliability for both in-person and remote wound assessment [67].

Deep learning–based diagnostic models achieved high accuracy in simultaneous wound segmentation and classification, indicating strong potential to improve efficiency and objectivity in DFU assessment, although their reliance on curated datasets may limit real-world applicability [36]. Building on this, explainability-integrated models not only improved segmentation performance but also enhanced interpretability, which is essential for clinical adoption, yet their usefulness depends on alignment with clinician reasoning processes [38]. In parallel, automated detection models demonstrated strong performance in identifying granulation tissue, supporting more objective evaluation of healing progression, although variability in wound characteristics across patients may affect consistency [37]. Furthermore, thermography-based segmentation frameworks improved detection accuracy by capturing physiological changes, but their dependence on specialized imaging modalities may limit scalability in resource-constrained settings [35]. While benchmark evaluations confirmed consistent performance across standardized datasets, these controlled environments may not fully reflect the heterogeneity of real-world clinical data [33]. This limitation is further reinforced by cross-dataset validation studies showing reduced performance on external datasets, highlighting persistent challenges in generalizability and the need for more diverse training data [34].

These advances suggest that AI-based measurement tools may help reduce subjectivity and improve consistency in DFU evaluation, although integration into routine workflows remains a key challenge. However, despite the strong technical performance reported in these studies, broader implementation will require additional validation across diverse patient populations and integration testing within clinical workflows.

XSL•FO
RenderX

Future research should focus on prospective validation, clinician feedback, and interoperability with electronic health records to ensure that AI-based DFU tools are not only accurate but also seamlessly embedded into routine wound care practices.

Diagnostic Classification and Condition Assessment

Recent advances in deep learning architectures further demonstrate improvements in diagnostic performance and model robustness. A hybrid deep learning architecture integrating convolutional networks, attention mechanisms, and transformer modules has shown high classification accuracy while improving interpretability through explainable AI techniques, suggesting enhanced capability for complex feature extraction [58]. An ensemble detection framework combining multiple object detection models demonstrated improved localization performance, supporting its potential use in automated screening [14]. In addition, a generative contrastive deep learning framework achieved very high classification accuracy and improved generalization across datasets, indicating continued methodological innovation in DFU image analysis [32]. Despite these developments, most studies primarily report technical performance without examining usability or integration into routine workflows. External validation across diverse patient populations also remains limited, highlighting the need for further prospective research before widespread clinical adoption. Despite these advances, most studies remain focused on technical validation, highlighting the ongoing need for external validation and real-world clinical evaluation.

Few-shot learning models achieved robust classification performance despite limited training data, suggesting strong potential for application in data-scarce environments, although their stability across heterogeneous clinical images remains uncertain [46]. Extending this, multimodal deep learning frameworks significantly improved classification accuracy by integrating complementary feature representations, yet the resulting increase in model complexity may hinder real-time clinical deployment [50]. Similarly, attention-based architectures enhanced feature extraction and classification performance, although issues related to interpretability and reproducibility across datasets remain insufficiently addressed [47]. Hybrid feature extraction approaches further improved classification accuracy by combining handcrafted and deep features, but their adaptability to new datasets may be limited [51]. In addition, CNN–vision transformer models achieved superior performance by capturing both local and global features, although their computational demands may restrict implementation in low-resource settings [45]. To address these constraints, efficient deep learning architectures achieved high classification accuracy while reducing computational complexity, suggesting improved feasibility for broader implementation [43]. Complementing this, lightweight models enabled accurate real-time DFU detection and grading, although potential trade-offs in handling complex cases require further investigation [49]. Beyond performance, explainable AI frameworks improved diagnostic transparency through interpretable visual outputs, yet their clinical relevance depends on clinician acceptance and understanding [48]. This trend is further supported by transformer-based explainable models, which enhance interpretability while maintaining high performance, although standardized evaluation of explainability remains lacking [44]. Finally, hybrid CNN-ELM approaches achieved competitive performance with reduced training complexity, offering efficiency advantages, but their scalability across larger and more diverse datasets remains uncertain [52].

Risk Prediction, Complication Detection, and Longitudinal Monitoring

Recent developments in AI-based models for DFU detection and risk prediction highlight their potential to support clinical decision-making and patient monitoring. Automated machine learning approaches have demonstrated promising performance compared with traditional methods, particularly for localized DFU risk assessment and real-time monitoring [27]. These approaches may offer more objective and consistent evaluations, which are critical for early intervention and personalized care planning. Additional evidence from recent studies further supports the potential role of AI-based classification models in contributing to clinical risk stratification. An AI-enhanced imaging framework combining segmentation and classification demonstrated strong performance in predicting ulcer severity based on Wagner grading, suggesting that image-based classification models may serve as indirect indicators of clinical risk and support early intervention strategies [42]. However, similar to earlier findings, these approaches primarily rely on retrospective datasets and require further prospective validation before being applied to recurrence prediction or long-term risk modeling.

Despite these promising results, broader implementation requires further validation and user-centered design improvements. Most studies focus on technical performance without addressing practical usability or integration into routine clinical workflows. Additionally, while internal validation shows strong metrics, external validation across diverse populations remains limited. Future work should prioritize prospective testing, compatibility with electronic health systems, and feedback from end users to ensure that AI tools are not only accurate but also clinically meaningful and scalable.

AI-based tools such as DFUCare, SPCD-based CNN models, and Siamese neural networks have been applied to support the monitoring of DFU progression and detection of complications including infection and ischemia [22,31]. These approaches may reduce variability associated with conventional visual assessment and support more consistent decision-making in complex DFU cases. Some studies also reported agreement between AI-generated recommendations and clinician judgment in amputation planning [56], suggesting a potential role for AI as a decision-support tool.

Emerging evidence also highlights the expanding role of AI in supporting longitudinal monitoring and remote management of DFU. An image analysis model integrating thermal and visual imaging demonstrated strong agreement with clinician assessments when evaluating wound healing trajectories, suggesting its potential to provide objective monitoring of wound progression [41]. An AI-enabled medical device designed for remote wound assessment showed high reliability in measuring wound parameters and classifying tissue characteristics, reinforcing its potential to support telemedicine

XSL•FO
RenderX

and reduce variability in clinical assessment [67]. These findings suggest a gradual shift from algorithm development toward clinically applicable monitoring tools, although broader implementation will require further validation across diverse clinical settings.

Machine learning models achieved high predictive accuracy in estimating minor amputation risk, highlighting their potential to support clinical decision-making, although reliance on retrospective datasets may limit predictive reliability in prospective settings [60]. Building on this, predictive models for hard-to-heal DFUs demonstrated strong discriminative ability, enabling earlier identification of patients at risk of poor outcomes, yet their generalizability across populations with different comorbidities remains uncertain [61]. Furthermore, externally validated models achieved reliable performance in predicting DFU infection, suggesting improved robustness, although broader validation across diverse health care systems is still required [62]. At a population level, predictive models achieved strong accuracy in identifying individuals at risk of developing DFU, supporting preventive strategies, but their integration into routine screening programs remains underexplored [63].

In addition to static risk prediction, temporal machine learning frameworks achieved high accuracy in predicting wound healing trajectories across clinical visits, enabling dynamic monitoring of patient outcomes, although their dependence on longitudinal data may limit use in settings with irregular follow-up [59]. Consistent with this, AI-based monitoring systems demonstrated strong agreement with clinical assessments in evaluating wound severity and progression, supporting their clinical applicability, yet their impact on workflow efficiency and patient outcomes remains insufficiently studied [64]. Moreover, thermography-based deep learning models improved early detection sensitivity for DFU-related complications, suggesting added diagnostic value, although limited access to thermal imaging may constrain widespread implementation [65]. Usability testing and workflow integration were rarely reported, highlighting the need for prospective, user-centered validation in clinical environments.

Clinical Decision Support and Automated Model Development

Mobile-based AI applications such as the Diabetic Foot Smart APP and eKare Insight have demonstrated potential to support more objective and standardized DFU measurement in both clinical and remote care settings [40,53]. These tools may help reduce interobserver variability and facilitate longitudinal documentation of wound-healing progress, addressing one of the main limitations of conventional visual assessment methods [17]. In addition, the use of automated image-based analysis enables clinicians to obtain more consistent wound measurements without requiring specialized equipment, which may be particularly beneficial in settings with limited access to wound-care expertise.

From an implementation perspective, mobile-based AI platforms align well with telemedicine and mobile health strategies, allowing wound assessment and monitoring to occur closer to the patient and reducing dependence on in-person visits [40,53]. Their compatibility with widely available smartphone devices increases accessibility and offers opportunities to support community-based care and remote clinical supervision.

Explainable AI approaches such as FusionNet aim to improve transparency in AI-supported DFU assessment by providing visual explanations of classifier outputs using techniques such as SHAP, LIME, or Grad-CAM [66]. These tools may help clinicians interpret AI-generated assessments alongside clinical indicators such as infection and ischemia status. Some studies reported comparable diagnostic performance between AI systems and clinicians [39]; however, validation remains limited and further evaluation is required across diverse clinical settings.

These findings highlight the growing role of explainable AI in enhancing clinical utility and decision-making alignment between AI systems and medical experts. Notably, FusionNet has demonstrated performance that surpasses junior and intermediate dermatologists while aligning closely with senior dermatologists, suggesting that AI can serve as a valuable support tool in DFU assessment [39]. This level of agreement is particularly relevant in critical care scenarios where early detection and accurate severity classification are essential for preventing complications such as amputation.

Moreover, the adoption of hardware-accelerated AI models like DFU_TFNet further enhances the practicality of these tools in real-time clinical environments [54]. By combining speed and precision, such models offer the potential for immediate deployment in primary care or telemedicine settings, especially in cases requiring rapid wound classification and triage.

The integration of explainable AI components within hybrid deep learning models further emphasizes the importance of transparency in AI-supported DFU assessment. A hybrid deep learning classification model incorporating attention mechanisms and interpretability frameworks provides visualization of decision pathways, which may enhance clinician understanding and trust in AI-generated outputs [58]. This growing emphasis on explainability reflects an important step toward improving clinical acceptability and facilitating integration into routine practice, although standardized evaluation of interpretability remains limited.

Despite these advantages, broader clinical deployment remains limited. Most studies primarily report technical feasibility and reliability, with relatively little evidence on real-world integration into existing clinical workflows, interoperability with electronic health information systems, or clinician and patient acceptability [17,40].

Machine learning–based clinical decision models achieved high predictive performance in supporting risk stratification and treatment planning, indicating strong potential to augment clinical decision-making, although their effectiveness ultimately depends on integration within existing clinical workflows [68]. Building on this, explainability-driven deep learning approaches improved transparency and clinician trust in AI systems, which is critical for adoption, yet the absence of standardized evaluation frameworks for interpretability remains a key barrier to widespread clinical implementation [44]. Furthermore, standardized evaluation frameworks and prospective multisite

XSL•FO
RenderX

validation are still required to determine the generalizability, safety, and long-term clinical value of these applications. Accordingly, the implementation of mobile AI tools for DFU assessment should continue to be approached as a supportive adjunct to clinical judgment rather than a replacement for comprehensive wound evaluation.

Comparison With Prior Work

This review extends prior work in wound assessment by focusing specifically on AI-based approaches for DFU evaluation. While earlier reviews emphasized manual or standardized clinical assessment tools, the present review highlights a shift toward automated systems designed to support diagnostic consistency and wound monitoring. AI-enabled platforms such as eKare Insight, FusionNet, and DFUCare illustrate this transition toward data-driven assessment and integration with digital health technologies.

Unlike traditional wound-care approaches, AI systems have the potential to reduce observer-dependent variability and support longitudinal tracking; however, translation into routine clinical practice remains limited by challenges in validation, usability, and workflow integration. This review therefore contributes by mapping current AI approaches, summarizing reported outcomes, and identifying key implementation gaps for future research.

Research Limitation

Lastly, although some AI tools reported agreement with clinician decision-making, only a small number of studies provided detailed statistical validation or external testing across different health care environments. As a result, the true robustness and reproducibility of these systems in everyday clinical practice remain uncertain. Many models were evaluated under controlled research conditions, which may not fully reflect the variability in patient characteristics, image quality, and workflow demands encountered in real-world settings. Future studies should therefore prioritize multicenter validation, standardized reporting, and prospective clinical evaluation to better determine the safety, reliability, and clinical readiness of AI-based DFU assessment tools. This highlights the need for rigorous, externally validated, and prospectively designed studies to establish clinical robustness and implementation feasibility.

Future Direction

Future research should prioritize the translation of AI-based DFU assessment tools into real-world clinical environments through prospective, multicenter evaluation. This includes external validation across diverse populations, imaging conditions, and care settings to improve generalizability and assess robustness beyond controlled research contexts. In addition, greater emphasis is needed on usability testing and user-centered design to ensure that AI systems align with clinician workflows, minimize cognitive burden, and enhance rather than replace clinical judgment.

Integration with digital health infrastructures such as electronic health records and telemedicine platforms represents another important direction to support continuity of care, remote monitoring, and multidisciplinary wound management. Explainable AI approaches also warrant further investigation to improve transparency and clinician trust, particularly in high-risk decision contexts. Finally, future work should adopt standardized reporting frameworks and evaluation metrics to enable consistent comparison across studies and facilitate the safe, scalable, and ethical implementation of AI-supported DFU assessment in clinical practice.

Ethical and Regulatory Implications of AI-Based DFU Assessment

The clinical application of AI-based DFU assessment tools must also be viewed through an ethical and regulatory lens. Key issues include transparency in algorithm development and decision pathways, fairness in performance across different patient groups, and protection of personal health data. AI models trained using limited or homogeneous datasets may introduce algorithmic bias, potentially resulting in unequal performance across demographic or clinical subgroups. Similarly, the use of clinical wound photographs raises important data governance and privacy concerns, particularly in telemedicine and mobile health environments. Ensuring model explainability, secure data handling, and appropriate regulatory oversight will therefore be essential to support clinician trust, safeguard patient rights, and promote equitable AI implementation in DFU care. Future research should incorporate ethical risk assessment alongside technical validation to support the responsible deployment of AI-enabled wound assessment systems.

Conclusions

The integration of AI into DFU assessment and management demonstrates promising technical and methodological advancements in areas such as diagnostic accuracy, risk prediction, wound monitoring, and clinical decision support. From machine learning to deep learning and explainable AI, current technologies offer innovative approaches that may reduce subjectivity and support early detection, particularly in resource-limited settings. However, the transition from experimental performance to routine clinical implementation remains hindered by challenges such as limited dataset diversity, lack of external validation, and insufficient integration into clinical workflows. To bridge this gap, future research should prioritize multicenter validation, user-centered design, interoperability, and alignment with clinical standards. Through such efforts, AI-based tools may become reliable and scalable components of DFU assessment in clinical practice.

Acknowledgments

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support (waiver) for the publication of this article. The authors used generative artificial intelligence tools: ChatGPT

XSL•FO
RenderX

[OpenAI] and Grammarly [Superhuman Platform Inc.] to assist in language refinement and editing of the manuscript. All content was critically reviewed and verified by the authors, who take full responsibility for the final version of the manuscript.

Funding

This study was supported by the Thematic Research Group (TRG) Batch 1, Hasanuddin University (Grant Number: 00518/UN4.22/PT.01.O3/2025). The funding body had no role in the design of the study; data collection, analysis, and interpretation; or in writing the manuscript.

Data Availability

All data analyzed in this study were obtained from previously published articles, which are cited within the manuscript. No new datasets were generated. The data supporting the findings of this review are available within the article and its supplementary materials, including Multimedia Appendix 1 (search strategy) and Checklist 1 (PRISMA-ScR checklist).

Authors' Contributions

Conceptualization: MZ, SY
Data curation: MZ, MJT, HB
Formal analysis: MZ
Investigation: MZ, MJT, HB
Methodology: MZ
Supervision: SY
Writing – original draft: MZ
Writing – review & editing: SY

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of literature search strategy: databases, keywords, and selection results.
[DOCX File, 13 KB - diabetes_v11i1e77925_app1.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 35 KB - diabetes_v11i1e77925_app2.docx]

References

1. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med 2017 Jun 15;376(24):2367-2375. [doi: 10.1056/NEJMra1615439] [Medline: 28614678]
2. Walsh JW, Hoffstad OJ, Sullivan MO, Margolis DJ. Association of diabetic foot ulcer and death in a population-based cohort from the United Kingdom. Diabet Med 2016 Nov;33(11):1493-1498. [doi: 10.1111/dme.13054] [Medline: 26666583]
3. Kooner SS, Sheehan B, Kendal JK, Johal H. Development of a simple multidisciplinary arthroplasty wound-assessment instrument: the SMArt Wound Tool. Can J Surg 2018 Oct 1;61(5):326-331. [doi: 10.1503/cjs.015017] [Medline: 30246994]
4. Palomo-López P, Losa-Iglesias ME, Becerro-de-Bengoa-Vallejo R, et al. Specific foot health-related quality-of-life impairment in patients with type II versus type I diabetes. Int Wound J 2019 Feb;16(1):47-51. [doi: 10.1111/iwj.12984] [Medline: 30168292]
5. Suva G, Sharma T, Campbell KE, Sibbald RG, An D, Woo K. Strategies to support pressure injury best practices by the inter-professional team: a systematic review. Int Wound J 2018 Aug;15(4):580-589. [doi: 10.1111/iwj.12901] [Medline: 29600545]
6. Houghton P. Wound assessment tools. : Wound Care Canada; 2018 URL: https://www.woundscanada.ca/docman/public/wound-care-canada-magazine/2018-16-no1/1273-wcc-summer-2018-v16n1-final-p-58-65-research-101/file [accessed 2026-06-13]
7. Bates-Jensen BM, McCreath HE, Harputlu D, Patlan A. Reliability of the Bates-Jensen wound assessment tool for pressure injury assessment: the pressure ulcer detection study. Wound Repair Regen 2019 Jul;27(4):386-395. [doi: 10.1111/wrr.12714] [Medline: 30828890]
8. Arisandi D, Oe M, Roselyne Yotsu R, et al. Evaluation of validity of the new diabetic foot ulcer assessment scale in Indonesia. Wound Repair Regen 2016 Sep;24(5):876-884. [doi: 10.1111/wrr.12464] [Medline: 27400025]
9. Oe M, Yotsu RR, Arisandi D, et al. Validity of DMIST for monitoring healing of diabetic foot ulcers. Wound Repair Regen 2020 Jul;28(4):539-546. [doi: 10.1111/wrr.12816] [Medline: 32281168]

XSL•FO
RenderX

10. Monteiro-Soares M, Martins-Mendes D, Vaz-Carneiro A, Sampaio S, Dinis-Ribeiro M. Classification systems for lower extremity amputation prediction in subjects with active diabetic foot ulcer: a systematic review and meta-analysis. Diabetes Metab Res Rev 2014 Oct;30(7):610-622. [doi: 10.1002/dmrr.2535] [Medline: 24523130]
11. Yu Z, Shen D, Jin Z, Huang J, Cai D, Hua XS. Progressive transfer learning. IEEE Trans Image Process 2022;31:1340-1348. [doi: 10.1109/TIP.2022.3141258] [Medline: 35025744]
12. Benbow M. Best practice in wound assessment. Nurs Stand 2016 Mar 2;30(27):40-47. [doi: 10.7748/ns.30.27.40.s45] [Medline: 26932649]
13. McCaughan D, Sheard L, Cullum N, Dumville J, Chetter I. Nurses' and surgeons' views and experiences of surgical wounds healing by secondary intention: a qualitative study. J Clin Nurs 2020 Jul;29(13-14):2557-2571. [doi: 10.1111/jocn.15279] [Medline: 32279371]
14. Wang Z, Tan X, Xue Y, et al. Smart diabetic foot ulcer scoring system. Sci Rep 2024 May 21;14(1):11588. [doi: 10.1038/s41598-024-62076-1] [Medline: 38773207]
15. Porter M. Developing wound services through digitising wound assessment: the benefits and challenges in a rural community nursing service. Br J Community Nurs 2023 Mar 2;28(Sup3):S20-S22. [doi: 10.12968/bjcn.2023.28.Sup3.S20] [Medline: 36809896]
16. Salam A, N A. Revolutionizing dermatology: the role of artificial intelligence in clinical practice. IJCED 2024 Jun 28;10(2):107-112. [doi: 10.18231/j.ijced.2024.021]
17. Chan KS, Chan YM, Tan AHM, et al. Clinical validation of an artificial intelligence-enabled wound imaging mobile application in diabetic foot ulcers. Int Wound J 2022 Jan;19(1):114-124. [doi: 10.1111/iwj.13603] [Medline: 33942998]
18. Barakat-Johnson M, Jones A, Burger M, et al. Reshaping wound care: evaluation of an artificial intelligence app to improve wound assessment and management. Stud Health Technol Inform 2024 Jan 25;310:941-945. [doi: 10.3233/SHTI231103] [Medline: 38269947]
19. Purnama AD, Yueniwati Y, Ismail D, et al. Screening diabetic foot ulcer using artificial intelligence modelling based on digital image analysis: a systematic review. ICON J 2025;10(1):117-134. [doi: 10.20956/icon.v10i1.43631]
20. Kabir MA, Samad S, Ahmed F, et al. Mobile apps for wound assessment and monitoring: limitations, advancements and opportunities. J Med Syst 2024 Aug 24;48(1):80. [doi: 10.1007/s10916-024-02091-x] [Medline: 39180710]
21. Goyal M, Yap MH, Reeves ND, Rajbhandari S, Spragg J. Fully convolutional networks for diabetic foot ulcer segmentation. In: 2017 IEEE International Conference on Systems, Man and Cybernetics (SMC): IEEE; 2017:618-623. [doi: 10.1109/SMC.2017.8122675]
22. Sendilraj V, Pilcher W, Choi D, et al. DFUCare: deep learning platform for diabetic foot ulcer detection, analysis, and monitoring. Front Endocrinol (Lausanne) 2024;15:1386613. [doi: 10.3389/fendo.2024.1386613] [Medline: 39381435]
23. Chan KS, Lo ZJ. Wound assessment, imaging and monitoring systems in diabetic foot ulcers: a systematic review. Int Wound J 2020 Dec;17(6):1909-1923. [doi: 10.1111/iwj.13481] [Medline: 32830440]
24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
25. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
27. Huang HN, Zhang T, Yang CT, et al. Image segmentation using transfer learning and Fast R-CNN for diabetic foot wound treatments. Front Public Health 2022;10:969846. [doi: 10.3389/fpubh.2022.969846] [Medline: 36203688]
28. Zhao X, Liu Z, Agu E, et al. Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network. IEEE Access 2019;7:179151-179162. [doi: 10.1109/access.2019.2959027] [Medline: 33777590]
29. Zhao N, Yu L, Fu X, et al. Application of a Diabetic Foot Smart APP in the measurement of diabetic foot ulcers. Int J Orthop Trauma Nurs 2024 Aug;54:101095. [doi: 10.1016/j.ijotn.2024.101095] [Medline: 38599150]
30. Wang L, Pedersen PC, Agu E, Strong D, Tulu B. Boundary determination of foot ulcer images by applying the associative hierarchical random field framework. J Med Imaging (Bellingham) 2019 Apr;6(2):024002. [doi: 10.1117/1.JMI.6.2.024002] [Medline: 31037245]
31. Aldoulah ZA, Malik H, Molyet R, Aljasem M. SwishRes-U-Net: a deep neural architecture for chronic wound segmentation. Biomed Signal Process Control 2025 Feb;100:107048. [doi: 10.1016/j.bspc.2024.107048]
32. Zoppo G, Marrone F, Pittarello M, et al. AI technology for remote clinical assessment and monitoring. J Wound Care 2020 Dec 2;29(12):692-706. [doi: 10.12968/jowc.2020.29.12.692] [Medline: 33320742]
33. Yap MH, Cassidy B, Byra M, et al. Diabetic foot ulcers segmentation challenge report: benchmark and analysis. Med Image Anal 2024 May;94:103153. [doi: 10.1016/j.media.2024.103153] [Medline: 38569380]
34. Lucho S, Naemi R, Castañeda B, Treuillet S. Can deep learning wound segmentation algorithms developed for a dataset be effective for another dataset? A specific focus on diabetic foot ulcers. IEEE Access 2024;12:173824-173835. [doi: 10.1109/ACCESS.2024.3502467]

XSL•FO
RenderX

35. Zaki W, Thanikachalam Y, Vajravelu A, Rathinam P, Thangavelu E. An integrated design for segmentation and classification of diabetic foot ulcers using thermography images. J Diabetes Metab Disord 2026 Jun;25(1):17. [doi: 10.1007/s40200-025-01813-3] [Medline: 41492273]
36. Zhou GX, Tao YK, Hou JZ, et al. Construction and validation of a deep learning-based diagnostic model for segmentation and classification of diabetic foot. Front Endocrinol (Lausanne) 2025;16:1543192. [doi: 10.3389/fendo.2025.1543192] [Medline: 40270716]
37. Lien ASY, Lai CY, Wei JD, Yang HM, Yeh JT, Tai HC. A granulation tissue detection model to track chronic wound healing in DM foot ulcers. Electronics (Basel) 2022;11(16):2617. [doi: 10.3390/electronics11162617]
38. Rathore PS, Kumar A, Nandal A, Dhaka A, Sharma AK. A feature explainability-based deep learning technique for diabetic foot ulcer identification. Sci Rep 2025 Feb 25;15(1):6758. [doi: 10.1038/s41598-025-90780-z] [Medline: 40000748]
39. Biswas S, Mostafiz R, Uddin MS, Paul BK. XAI-FusionNet: diabetic foot ulcer detection based on multi-scale feature fusion with explainable artificial intelligence. Heliyon 2024 May 30;10(10):e31228. [doi: 10.1016/j.heliyon.2024.e31228] [Medline: 38803883]
40. Fadhel MA, Alzubaidi L, Gu Y, Santamaría J, Duan Y. Real-time diabetic foot ulcer classification based on deep learning & parallel hardware computational tools. Multimed Tools Appl 2024;83(27):70369-70394. [doi: 10.1007/s11042-024-18304-x]
41. Shandilya G, Gupta S, Gupta D, et al. Proposed dense variational autoencoder model integrated with contrastive learning for foot ulcer classification. Sci Rep 2026 Jan 2;16(1):4171. [doi: 10.1038/s41598-025-34225-7] [Medline: 41484318]
42. Sarmun R, Chowdhury MEH, Murugappan M, et al. Diabetic foot ulcer detection: combining deep learning models for improved localization. Cogn Comput 2024 May;16(3):1413-1431. [doi: 10.1007/s12559-024-10267-3]
43. Almufadi NF, Alhasson HF, Alharbi SS. E-DFu-Net: an efficient deep convolutional neural network models for diabetic foot ulcer classification. Biomol Biomed 2025 Jan 14;25(2):445-460. [doi: 10.17305/bb.2024.11117] [Medline: 39319898]
44. Karthik R, Ajay A, Jhalani A, Ballari K, K S. An explainable deep learning model for diabetic foot ulcer classification using Swin transformer and efficient multi-scale attention-driven network. Sci Rep 2025 Feb 3;15(1):4057. [doi: 10.1038/s41598-025-87519-1] [Medline: 39900977]
45. Sait ARW, Nagaraj R. Diabetic foot ulcers detection model using a hybrid convolutional neural networks-vision transformers. Diagnostics (Basel) 2025 Mar 15;15(6):736. [doi: 10.3390/diagnostics15060736] [Medline: 40150079]
46. Wang C, Yu Z, Long Z, Zhao H, Wang Z. A few-shot diabetes foot ulcer image classification method based on deep ResNet and transfer learning. Sci Rep 2024 Dec 2;14(1):29877. [doi: 10.1038/s41598-024-80691-w] [Medline: 39622873]
47. Ajay A, Singh Bisht A, Karthik R. Dense-ShuffleGCANet: an attention-driven deep learning approach for diabetic foot ulcer classification using refined spatio-dimensional features. IEEE Access 2025;13:5507-5521. [doi: 10.1109/ACCESS.2024.3524549]
48. Mahmud MI, Reza MS, Akash MOA, Elias F, Ahmed N. DFU_DIALNet: towards reliable and trustworthy diabetic foot ulcer detection with synergistic confluence of Grad-CAM and LIME. PLoS ONE 2025;20(9):e0330669. [doi: 10.1371/journal.pone.0330669] [Medline: 40892859]
49. Girmaw DW, Taye GB. MobileNetV2 model for detecting and grading diabetic foot ulcer. Discov Appl Sci 2025;7(4). [doi: 10.1007/s42452-025-06745-4]
50. Salur MU. A multimodal dual-stream cross-attention deep learning framework for diabetic foot ulcer classification. Applied Sciences 2026;16(4):1993. [doi: 10.3390/app16041993]
51. Al-Garaawi N, Ebsim R, Alharan AFH, Yap MH. Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks. Comput Biol Med 2022 Jan;140:105055. [doi: 10.1016/j.compbiomed.2021.105055] [Medline: 34839183]
52. Arnia F, Saddami K, Roslidar R, Muharar R, Munadi K. Towards accurate diabetic foot ulcer image classification: leveraging CNN pre-trained features and extreme learning machine. Smart Health (2014) 2024 Sep;33:100502. [doi: 10.1016/j.smhl.2024.100502]
53. Xie P, Li Y, Deng B, et al. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int Wound J 2022 May;19(4):910-918. [doi: 10.1111/iwj.13691] [Medline: 34520110]
54. Xiaoling W, Shengmei Z, BingQian W, et al. Enhancing diabetic foot ulcer prediction with machine learning: a focus on localized examinations. Heliyon 2024 Oct 15;10(19):e37635. [doi: 10.1016/j.heliyon.2024.e37635] [Medline: 39386877]
55. Goyal M, Reeves ND, Rajbhandari S, Ahmad N, Wang C, Yap MH. Recognition of ischaemia and infection in diabetic foot ulcers: dataset and techniques. Comput Biol Med 2020 Feb;117:103616. [doi: 10.1016/j.compbiomed.2020.103616] [Medline: 32072964]
56. Toofanee MSA, Dowlut S, Hamroun M, et al. DFU-Helper: an innovative framework for longitudinal diabetic foot ulcer diseases evaluation using deep learning. Appl Sci 2023;13(18):10310. [doi: 10.3390/app131810310]
57. Bhattacharya T, Chakraborty S, Goyal G, Singh M, Jude BE, Mukherjee S. AI-enhanced imaging for diabetic foot ulcer risk assessment and diagnosis: a retrospective cohort study. J Diabetes Sci Technol 2026 Jan 15:19322968251409761. [doi: 10.1177/19322968251409761] [Medline: 41540796]
58. Sharma N, Mirza S, Rastogi A, Mahapatra PK. Utilizing mask R-CNN for automated evaluation of diabetic foot ulcer healing trajectories: a novel approach. TS 2023 Aug 31;40(4):1601-1610. [doi: 10.18280/ts.400428]

XSL•FO
RenderX

59. Basiri R, Saleh A, Khan SS, Popovic MR. Temporal machine learning framework for diabetic foot ulcer healing trajectory prediction. Biomed Eng Online 2026 Feb 5;25(1):41. [doi: 10.1186/s12938-026-01529-2] [Medline: 41645209]
60. Wang S, Wang J, Zhu MX, Tan Q. Machine learning for the prediction of minor amputation in University of Texas grade 3 diabetic foot ulcers. PLoS ONE 2022;17(12):e0278445. [doi: 10.1371/journal.pone.0278445] [Medline: 36472981]
61. Wang S, Xia C, Zheng Q, Wang A, Tan Q. Machine learning models for predicting the risk of hard-to-heal diabetic foot ulcers in a Chinese population. Diabetes Metab Syndr Obes 2022;15:3347-3359. [doi: 10.2147/DMSO.S383960] [Medline: 36341229]
62. Nie X, Jiang Y, Meng X, et al. Development and external validation of a machine learning model for predicting wound infection in diabetic foot ulcers. Diabetes Metab Syndr Obes 2026;19:586810. [doi: 10.2147/DMSO.S586810] [Medline: 41821605]
63. Zhang Y, Tian Y, Jian Y, et al. Development and external validation of machine-learning based models to predict diabetic foot ulcer in diabetes population. Front Endocrinol (Lausanne) 2025;16:1692917. [doi: 10.3389/fendo.2025.1692917] [Medline: 41473241]
64. Satheesh SS, Rayampalli A, Prabhune AG, Sri Hari VR. Development and validation of a predictive AI framework for diabetic foot ulcer monitoring and severity assessment: a step towards self-monitoring and primary care integration. Healthc Inform Res 2026 Jan;32(1):69-76. [doi: 10.4258/hir.2026.32.1.69] [Medline: 41681001]
65. Eldin AS, Ahmoud AS, Hamza HM, Ardah H. Enhancing early detection of diabetic foot ulcers using deep neural networks. Diagnostics (Basel) 2025 Aug 9;15(16):1996. [doi: 10.3390/diagnostics15161996] [Medline: 40870850]
66. Mert M, Vahabi A, Daştan AE, et al. Artificial intelligence's suggestions for level of amputation in diabetic foot ulcers are highly correlated with those of clinicians, only with exception of hindfoot amputations. Int Wound J 2024 Oct;21(10):e70055. [doi: 10.1111/iwj.70055] [Medline: 39353602]
67. Rifat RA, Bhoyan FH, Mehedi MHK, Hossain MK, Hossen MJ, Mridha MF. ConMatFormer: a multi-attention and transformer integrated ConvNext based deep learning model for enhanced diabetic foot ulcer classification. Results in Engineering 2025 Dec;28:108248. [doi: 10.1016/j.rineng.2025.108248]
68. Gao L, Liu Z, Han S, Wang J. A machine-learning-based clinical decision model for predicting amputation risk in patients with diabetic foot ulcers: diagnostic performance and practical implications. Diagnostics (Basel) 2025 Dec 10;15(24):3142. [doi: 10.3390/diagnostics15243142] [Medline: 41464143]
69. Almufadi N, Alhasson HF. Classification of diabetic foot ulcers from images using machine learning approach. Diagnostics (Basel) 2024 Aug 19;14(16):1807. [doi: 10.3390/diagnostics14161807] [Medline: 39202295]

Abbreviations

AHRF: associative hierarchical random field
AI: artificial intelligence
CNN: convolutional neural network
DEPA: depth, extent of bacterial colonization, phase of healing, and associated aetiology
DFU: diabetic foot ulcer
DFUAS: Diabetic Foot Ulcer Assessment Scale
DFUC: Diabetic Foot Ulcer Challenge
DMIST: depth, maceration, inflammation/infection, size, tissue type, type of wound edge, tunneling
ELM: extreme learning machine
FPGA: field-programmable gate array
GPU: graphics processing unit
Grad-CAM: gradient-weighted class activation mapping
IoU: intersection over union
LiDAR: light detection and ranging
LightGBM: light gradient boosting machine
LIME: local interpretable model-agnostic explanations
PCC: population, concept, context
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SHAP: Shapley Additive Explanations
SPCD: superpixel color descriptor

XSL•FO
RenderX

Edited by A Stone, N Cahill; submitted 22.May.2025; peer-reviewed by K Kuppan, ME Heidari, Z Wang; revised version received 29.Apr.2026; accepted 29.Apr.2026; published 08.Jul.2026.

Please cite as:
Zulfiqar M, Yusuf S, Taming MJ, Burhan H
AI-Based Models for Diabetic Foot Ulcer Assessment: Scoping Review
JMIR Diabetes 2026;11:e77925
URL: https://diabetes.jmir.org/2026/1/e77925
doi:10.2196/77925

© Muhamad Zulfiqar, Saldy Yusuf, Muhammad Jufri Taming, Herlina Burhan. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 8.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Twin Applications in Diabetes Management: Scoping Review

Fatemeh Sarani Rad[1*], MSc; Maryam Jafarpour[2*], PhD; Ehsan Bitaraf[3*], MD; Katayoon Khaleghdadi[4*], BS; Juan Li[1*], PhD

[1]Computer Science Department, North Dakota State University, 1320 Albrecht Blvd, Fargo, ND, United States

[2]Institute of Outcomes Research, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria

[3]Center for Technology and Innovation in Cardiovascular Informatics, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran

[4]Department of Computer Engineering, Islamic Azad University, Mashhad Branch, Mashhad, Iran

[*]all authors contributed equally

Corresponding Author:
Juan Li, PhD
Computer Science Department, North Dakota State University, 1320 Albrecht Blvd, Fargo, ND, United States

Abstract

Background: Digital twin (DT) systems have emerged as a promising approach in health care, enabling real-time, patient-specific virtual modeling and personalized interventions. In diabetes care, DTs offer the potential to revolutionize glucose management, decision support, and therapy personalization through integration of real-time and longitudinal patient data.

Objective: This scoping review mapped the current landscape of DT applications in diabetes and synthesized evidence across 13 research questions organized into 7 thematic domains: system design, target conditions, data sources, personalization strategies, intelligence and adaptability, validation methods, and implementation considerations.

Methods: This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and JBI methodological guidance for scoping reviews. A literature search was performed in PubMed, IEEE Xplore, Scopus, and Web of Science for studies published up to April 2025; all databases were last searched on June 23, 2025. Eligible studies were original empirical articles in English that described patient-specific DT systems or closely related individualized virtual models applied to diabetes diagnosis, monitoring, management, treatment, or complication-related care. Reviews, editorials, commentaries, theoretical papers without original data, and studies not focused on diabetes were excluded. Furthermore, FSR, MJ, and KK independently screened records and assessed full texts, with disagreements resolved through discussion and, when needed, by EB. Data were charted using a structured framework based on 13 predefined research questions, and were synthesized descriptively and thematically.

Results: Of 208 records identified, 123 underwent title and abstract screening, 39 full texts were assessed for eligibility, and 28 studies were included. Most studies focused on type 1 or type 2 diabetes and used data-driven, hybrid, or simulation-based DT approaches. Common clinical applications included therapeutic control, glucose prediction, decision support, and disease management. Lifestyle data, wearables, continuous glucose monitoring, and electronic health records were the dominant inputs, while personalization relied on adaptive feedback, insulin optimization, and behavior-driven tools. Intelligent features, such as adaptive learning, explainable artificial intelligence, and real-time synchronization, enhanced adaptability, although human oversight was rare. Validation was mainly retrospective or simulation-based, with few clinical trials; reported outcomes included improved hemoglobin A_{1c}, time-in-range, and reduced hypoglycemia. Ethical discussions focused on data privacy, while implementation barriers centered on validation gaps, data quality, and workflow integration.

Conclusions: DT research in diabetes is expanding and shows strong potential for personalized and data-driven care; however, the evidence base remains heterogeneous, inconsistently reported, and limited in prospective clinical validation. Key gaps include standardized definitions, robust real-world evaluation, fairness and governance considerations, and integration into clinical workflows. Future work should prioritize clinically grounded validation, regulatory readiness, and interoperable architectures to support safe, equitable, and scalable implementation.

(JMIR Diabetes 2026;11:e83059) doi:10.2196/83059

KEYWORDS

digital twin; diabetes mellitus; continuous glucose monitoring; automated insulin delivery; machine learning; clinical decision support; ethics

XSL•FO
RenderX

Introduction

A digital twin (DT) is a dynamic, virtual representation of a physical system—such as a patient—that is continuously updated with real-world data and computational models to support prediction, simulation, and decision-making [1]. In health care, DTs are a powerful tool for personalized medicine, providing real-time, data-driven insights tailored to individual patients [2,3].

Diabetes mellitus, encompassing both type 1 and type 2 diabetes, remains a major chronic health condition requiring highly individualized care [4,5]. The complexity of diabetes management—driven by variability in disease trajectories, treatment responses, and complication risks—requires approaches that move beyond traditional one-size-fits-all models. DTs address this need by simulating glycemic dynamics, forecasting outcomes, and supporting therapy optimization on a patient-specific basis [2,3,6,7]. These models integrate diverse data sources, such as continuous glucose monitoring (CGM), insulin dosing records, electronic health records (EHRs), wearable sensors, genomic information, and lifestyle factors [3,6,7].

Recent research highlights the potential of DTs in diabetes for applications, such as predicting disease progression, personalizing nutrition, enhancing automated insulin delivery systems, and supporting self-management [3,7-11]. For instance, DT frameworks that combine machine learning, multimodal data, and mechanistic modeling have been used to predict glycemic and complication-related outcomes in diabetes [3,7,8,12,13]. Early-phase clinical and real-world studies suggest potential improvements in glycemic control, reduced medication use, and enhanced metabolic outcomes with DT-based interventions [8,11,14-18].

However, several barriers still hinder broader adoption and clinical integration. Key challenges include data integration and model personalization [3,6,7], limited interoperability across devices and systems [2,6,7], the absence of standardized validation and regulatory pathways [2,6,7], and unresolved concerns around data privacy and ethical use [2,6].

Despite promising progress, DT research in diabetes remains fragmented and undervalidated. While some reviews have examined digital health tools in diabetes or explored DTs in general health care contexts [6], no previous review has systematically synthesized DT applications in diabetes across key dimensions such as system design, personalization, data integration, validation, and implementation. This gap limits the ability of researchers, clinicians, and developers to assess maturity levels, identify best practices, and guide future development.

To address this gap, we conducted a scoping review guided by the following research questions. The review addresses 13 research questions (RQs) grouped under 7 thematic domains to improve clarity and synthesis.

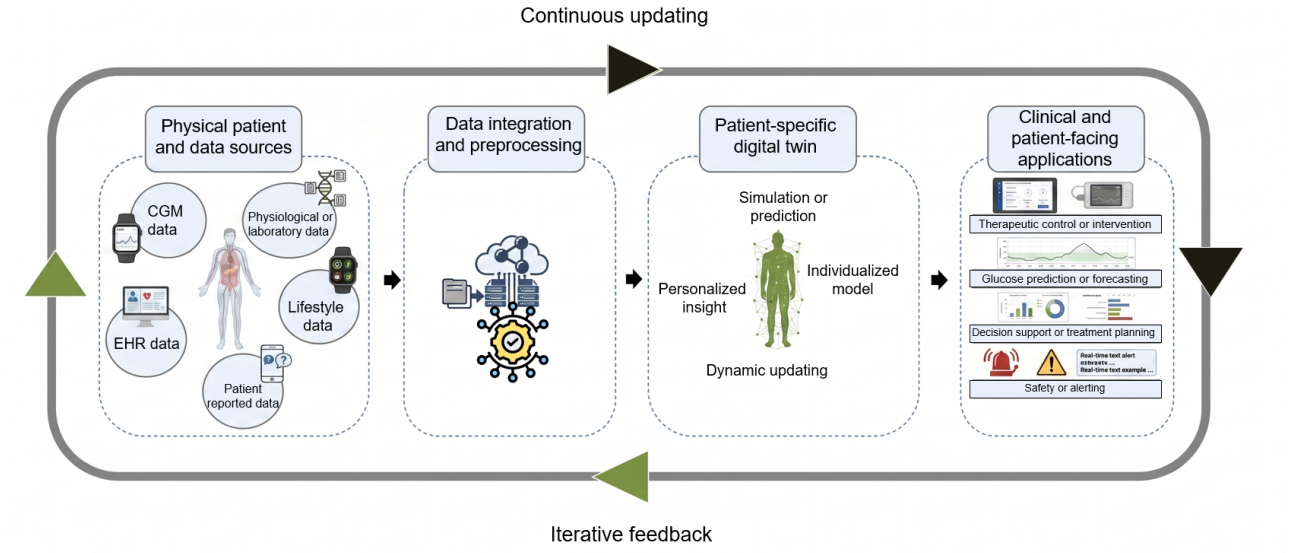
1. System design and modeling foundations:
   - RQ1: What types of DT models have been developed for diabetes care and management?
   - RQ2: What system components are included in these models?
   - RQ3: What modeling approaches are used in these systems?
2. Target conditions and use context:
   - RQ4: What types of diabetes are addressed by these DT applications?
   - RQ5: What clinical goals do these DTs aim to support?
3. Data sources and personalization mechanisms:
   - RQ6: What data sources are used to build or update DTs for diabetes?
   - RQ7: How are DTs used to enable personalized care or self-management in diabetes?
4. Intelligence and adaptability:
   - RQ8: How do the DTs handle uncertainty, real-time data updates, and model interpretability?
5. Evaluation and validation:
   - RQ9: What outcomes have been reported from applying DTs in diabetes care?
   - RQ10: What methods have been used to validate these DT systems?
6. Implementation and governance:
   - RQ11: What ethical or legal issues are raised regarding the use of DTs in diabetes care?
   - RQ12: What barriers and enablers are reported for implementing DT systems in clinical practice?
7. Research and development gaps:
   - RQ13: What gaps in knowledge or practice are identified in the literature on DTs in diabetes?

By systematically synthesizing evidence across these domains, this review provides a comprehensive overview of the current state of DT research in diabetes. The findings aim to inform researchers, clinicians, and technology developers about prevailing trends, methodological practices, and future opportunities for advancing personalized diabetes care through DT technologies [2,4,5].

Figure 1 presents a synthesized architecture of DT systems in diabetes based on the common components identified across the included studies.

XSL•FO
RenderX

Figure 1. Synthesized architecture of digital twin systems in diabetes based on the included studies. Common components included multimodal patient data inputs, data integration and preprocessing, a patient-specific digital twin core, and clinical or patient-facing applications, such as therapeutic control, glucose prediction, decision support, treatment planning, and safety alerting. The outer loop represents continuous updating of the digital twin using incoming patient data and iterative feedback to support adaptive personalization. CGM: continuous glucose monitoring; EHR: electronic health record.



Methods

Overview

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and JBI methodological guidance for scoping reviews [19,20].

Information Sources and Search Strategy

A comprehensive literature search was conducted through PubMed, IEEE Xplore, Scopus, and Web of Science. Studies published up to April 2025 were considered, and all databases were last searched on June 23, 2025. The search strategy combined terms related to "digital twin," "diabetes," and "healthcare" using Boolean operators. The detailed search strategy is provided in the Multimedia Appendix 1. Reference lists of included studies and relevant reviews were also manually screened to identify additional records.

The search strategies developed for PubMed, Web of Science, IEEE Xplore, and Scopus were imported into the Triple-A (Article Analysis Assistant) software [21]. The tool was used to integrate bibliographic metadata, automatically remove duplicate records based on DOI, and perform additional deduplication using title, publication year, and author names. Reviewer decisions were subsequently imported into the platform, and the finalized dataset was prepared for downstream analysis and thematic synthesis.

Eligibility Criteria

Eligibility criteria were established a priori to ensure consistency and reproducibility during screening.

Inclusion Criteria

Studies were included if they were original empirical research articles, including peer-reviewed journal papers, conference proceedings, or preprints. Studies were eligible if they reported on the development, validation, implementation, or clinical evaluation of DT systems for diabetes, including type 1, type 2, gestational, or related complications. Research involving patient-specific modeling, simulation, or data-driven approaches relevant to diabetes diagnosis, management, or treatment was included. Articles addressing applications in personalized or precision medicine, clinical decision support, or individualized therapy for diabetes were also included. Publications were required to be written in English, with a structured abstract and an accessible full text.

Exclusion Criteria

Studies were excluded if they were review articles, meta-analyses, editorials, commentaries, book chapters, or theoretical or conceptual papers without original data. Studies focused on DTs for diseases or systems other than diabetes, such as cardiovascular, neurological, or orthopedic applications, were excluded. Articles lacking an abstract or full text, or published in languages other than English, were also excluded.

These criteria were set before the screening process to maintain consistency and transparency in study selection. During screening, FSR, MJ, and KK independently assessed each record for eligibility using the predefined criteria. Discrepancies or uncertainties were resolved through discussion, with EB consulted when necessary.

Study Selection

The selection process involved 3 stages—identification, screening, and eligibility assessment. We initially identified 208 studies from 4 major databases—PubMed (n=47), IEEE Xplore (n=1), Scopus (n=107), and Web of Science (n=53). During identification, 85 articles were excluded due to duplication, lack of an abstract, absence of original data, or being published in a language other than English.

Following this step, 123 articles proceeded to screening. At the screening stage, 84 articles were excluded according to the

XSL•FO
RenderX

predetermined exclusion criteria. As a result, 39 articles advanced to eligibility assessment, and 28 were included in the final review [3,7-18,22-36]. The 11 full-text articles excluded at the eligibility stage and the reasons for exclusion are listed in Multimedia Appendix 2.

Screening was conducted in 2 stages:

- Title and abstract screening: FSR, MJ, and KK independently assessed each record against the predefined eligibility criteria.
- Full-text screening: Articles passing the first stage were retrieved in full and assessed for final inclusion.

To ensure consistent inclusion decisions, the following screening questions were applied, reflecting the key characteristics of DTs and their application in diabetes care (Table 1). For the purposes of this review, a study was considered to describe a DT if it included a patient-specific virtual representation or individualized computational model linked to diabetes-related data and intended for prediction, simulation, monitoring, or decision support. Studies using terms such as "virtual patient" or "simulation model" were included only if these DT-defining characteristics were present. Generic population-level models without individualized representation or diabetes-specific application were excluded.

Table . Filtering questions for study selection.

Screening question	Decision criteria
FQ1[a]: Does the study discuss or apply DT[b] technology?	Include only if the study explicitly referred to a DT or described a patient-specific virtual representation or individualized computational model linked to diabetes-related data and intended for prediction, simulation, monitoring, or decision support.
FQ2: Is the study focused on diabetes or diabetes-related conditions?	Include only if the main population or application domain involves diabetes (type 1, type 2, and gestational) or closely related metabolic conditions (eg, diabetic nephropathy and retinopathy).
FQ3: Is the DT model tailored to individual patients or based on patient-specific data?	Include only if the DT system is personalized using real or simulated patient-specific data (eg, glucose levels, insulin history, CGM[c], and EHRs[d]). Exclude if the system is generic or population-level only.

[a]FQ: filtering question.

[b]DT: digital twin.

[c]CGM: continuous glucose monitoring.

[d]EHR: electronic health record.

Data Extraction and Thematic Framework

Data from the 28 included studies were charted using a structured framework guided by the 13 predefined research questions introduced in the Introduction section. These research questions were organized into seven thematic domains to facilitate systematic synthesis: (1) system design and modeling foundations (RQ1, RQ2, and RQ3), (2) target conditions and use context (RQ4 and RQ5), (3) data sources and personalization mechanisms (RQ6 and RQ7), (4) intelligence and adaptability (RQ8), (5) evaluation and validation (RQ9 and RQ10), (6) implementation and governance (RQ11 and RQ12), and (7) research and development gaps (RQ13).

Each included study was analyzed systematically using this framework. Categories were not mutually exclusive, and individual studies could be charted under more than 1 category where appropriate. The full study characteristics and data charting table are provided in Multimedia Appendix 3.

Consistent with scoping review methodology, formal risk-of-bias, reporting bias, and certainty-of-evidence assessments were not performed because the aim was to map the breadth, characteristics, and gaps in a heterogeneous body of literature rather than to compare intervention effects or generate pooled estimates.

A review protocol and project materials for this scoping review were made available through the Open Science Framework (OSF) [37].

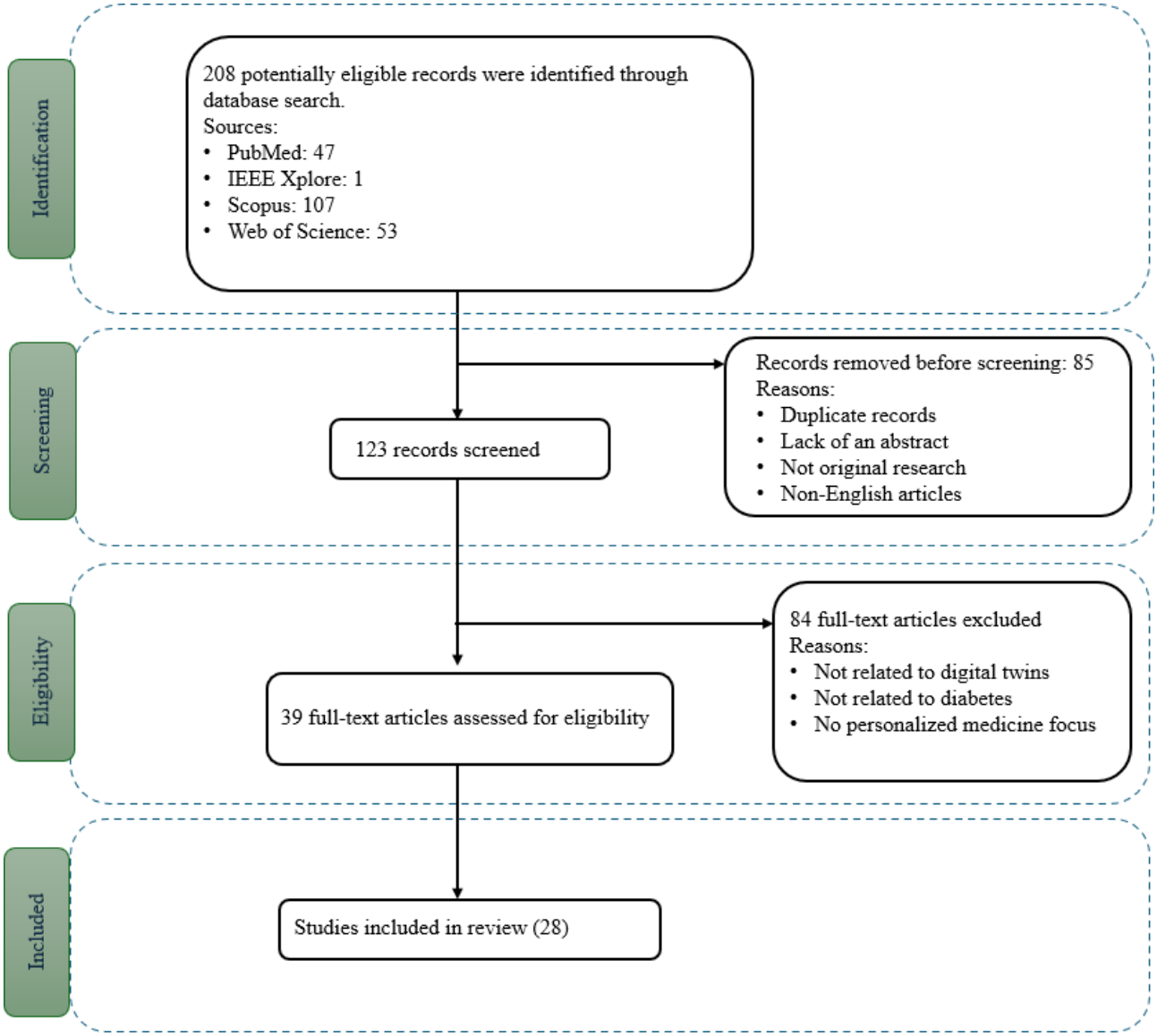
Results

Overview

Across the 28 included studies (as shown in the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flow diagram in Figure 2) [3,7-18,22-36], DT systems for diabetes exhibited diverse architectures, data sources, and application goals. Most models were data-driven or hybrid (artificial intelligence [AI]+mechanistic), while purely mechanistic and conceptual designs were less common. Core system components included machine learning (ML) or AI modules, decision support layers, and real-time simulation engines. The majority of DTs leveraged CGM, wearables, and lifestyle data, with increasing use of patient-specific models to enable personalized therapy, behavioral nudges, and simulation-based feedback.

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram of study selection process.



ML was the dominant modeling approach, while reinforcement learning, control theory, and signal processing appeared less frequently. Strategies for uncertainty management and interpretability were adopted inconsistently, with adaptive learning and explainable AI used in some studies, but with limited human-in-the-loop oversight. Reported outcomes most often focused on glycemic control (eg, hemoglobin A_{1c} [HbA_{1c}], time-in-range [TIR], and reduced hypoglycemia), alongside improvements in predictive accuracy, metabolic markers, and patient engagement. However, external clinical validation remained scarce, with most evaluations based on retrospective datasets or simulations.

Ethical considerations—mainly privacy and transparency, with occasional references to accountability and bias—were inconsistently addressed. Implementation barriers included validation limitations, data quality issues, model limitations, and workflow misalignment. Finally, the literature highlights persistent research gaps in integration with real-world systems, scalability, and methodological rigor that must be addressed to advance DT systems into clinical use.

System Design and Modeling Foundations (RQ1, RQ2, RQ3)

Overview

This section describes how DT models in diabetes are structured and modeled. It summarizes the types of models used (RQ1), the core system components included (RQ2), and the computational modeling strategies adopted (RQ3). Together, these questions cover the architectural and technical foundations of DTs in diabetes.

Model Types (RQ1)

DT models in diabetes care fall into 4 main categories—data-driven, hybrid, mechanistic, and conceptual. Data-driven models—most commonly using ML or deep learning—were used in half of the studies and focused on prediction and classification tasks. Hybrid models, which combine physiological modeling with AI, support real-time control systems, such as automated insulin delivery. Mechanistic models appeared less frequently and were primarily used in simulation studies. Conceptual frameworks were rare and largely theoretical. Table 2 summarizes the types of DT models reported

XSL•FO
RenderX

in diabetes care, with representative examples from included studies.

Table . Types of digital twin models used in diabetes care with representative examples from included studies (research question 1). Percentages may exceed 100% because individual studies could be coded into more than one category.

Model type	Key characteristics	Studies, n	Representative examples
Data-driven	ML[a], DL[b], RL[c]; CGM[d]-based	14	Shamanna et al [8], Shamanna et al [14], Vaskovsky and Chvanova [22], Shamanna et al [15]
Hybrid	ML+mechanistic model	10	Sarani Rad et al [3], Cappon et al [9], Colmegna et al [23], Ahmadasas et al [10]
Mechanistic	ODEs[e], simulations	5	Young et al [24], Thamotharan et al [25], Wang et al [26], Zavitsanou et al [27]
Conceptual	Framework only	1	Mishra et al [28]

[a]ML: machine learning.

[b]DL: deep learning.

[c]RL: reinforcement learning.

[d]CGM: continuous glucose monitoring.

[e]ODE: ordinary differential equation.

Key findings included:

- Data-driven models (14 studies, 50%): Applied for HbA_{1c} forecasting, glycemic risk scoring, and behavior modeling [8,11-18,22,31,32,34,35].
- Hybrid models (10 studies, 35.7%): Enabled adaptive insulin dosing and feedback control by integrating ML with mechanistic physiology [3,7-10,23,29,30,33,36].
- Mechanistic models (5 studies, 17.9%): Focused on ODE-based glucose-insulin dynamics for simulation and metabolic exploration [24-28].
- Conceptual frameworks (1 study, 3.6%): Proposed theoretical DT architecture without implementation [28].

System Components (RQ2)

Most DT systems consisted of modular components supporting prediction, simulation, control, and user interaction. The most common modules were ML or AI components, followed by simulation engines and data integration layers. User-facing dashboards and decision support or control modules were also frequently described, while personalization layers, backend infrastructure, and rule-based systems were less common. Table 3 summarizes the system component categories reported in diabetes DT models, including their functions and representative examples.

XSL•FO
RenderX

Table . System component categories in diabetes digital twin models with representative examples from included studies (research question 2). Percentages may exceed 100% because individual studies could be coded into more than one category.

System component category	Key characteristics	Studies, n	Representative examples
ML[a]/AI[b] module	LSTM[c], CNN[d], reinforcement learning	20	Zhang et al [7], Pellizzari et al [29], Chen et al [30], Joshi et al [11]
Simulation engine	Glucose-insulin model, ReplayBG engine, ODE[e]-based simulator	17	Wang et al [26], Zavitsanou et al [27], Mishra et al [28], Leszczełowska et al [31]
Data integration layer	CGM[f] devices, IoT[g] sensors, preprocessing layer	12	Vaskovsky and Chvanova [22], Ahmadasas et al [10], Villa-Tamayo et al [32], Shamanna et al [16]
User interface or dashboard	Mobile apps, web dashboards, patient interfaces	11	Colmegna et al [23], Leszczełowska et al [31], Shamanna et al [17], Shamanna et al [18]
Decision support or control feedback module	MPC[h], PID[i] controller, feedback system	10	Shamanna et al [8], Cappon et al [9], Young et al [24], Zhu et al [33]
Intervention or recommendation engine	GPT[j]-based module, precision nutrition, lifestyle recommendations	9	Sarani Rad et al [3], Shamanna et al [15], Young et al [24], Shamanna et al [17]
Personalization layer	Personalization engine, patient-specific tuning	5	Cappon et al [9], Young et al [24], Pellizzari et al [29], Chen et al [30]
Monitoring and alerts	Real-time alerts, patient monitoring, CGM-based tracking	4	Shamanna et al [15], Shamanna et al [16], Shamanna et al [18], Vaskovsky et al [34]
Backend or platform infrastructure	Cloud platform, database engine, analytics engine	3	Vaskovsky et al [34], Chahal et al [35], Cappon et al [36]
Knowledge representation or semantic layer	Knowledge graphs, ontologies	2	Sarani Rad et al [3], Zhang et al [7]
Rule-based decision system	Expert system, rule tables	2	Shamanna et al [8], Zhu et al [33]

[a]ML: machine learning.
[b]AI: artificial intelligence.
[c]LSTM: long short-term memory.
[d]CNN: convolutional neural network.
[e]ODE: ordinary differential equation.
[f]CGM: continuous glucose monitoring.
[g]IoT: internet of things.
[h]MPC: model predictive control___.
[i]PID: proportional-integral-derivative___.
[j]GPT: generative pre-trained transformer.

Key findings included:

- ML or AI modules (20, 71.4% studies) were central to prediction, therapy optimization, and personalization [7-14,16-18,22,23,29-35].
- Simulation engines (17, 60.7% studies) provided physiological modeling and glucose-insulin dynamics for testing and validation [3,7,9,13,15,17,18,24-31,33,36].
- Data integration layers (12, 42.9% studies) supported real-time data collection from CGM, internet-of-things sensors, and preprocessing pipelines [3,8,10,13,14,16,22,23,32,34-36].
- User interfaces (11, 39.3% studies) enabled interaction for patients and clinicians through mobile apps or dashboards [8,13,14,16-18,23,25,31,32,36].

Modeling Approaches (RQ3)

The computational strategies used in diabetes DT systems reflect both the predictive and control needs of these models. While ML was the dominant method, several studies incorporated reinforcement learning, control theory, and signal processing for adaptive and real-time decision-making. Table 4 summarizes the range of modeling techniques reported across included studies, with representative examples.

XSL•FO
RenderX

Table . Modeling techniques and approaches used in diabetes digital twin systems, with representative examples from included studies (research question 3). Percentages may exceed 100% because individual studies could be coded into more than one category.

Modeling approach category	Key characteristics	Studies, n	Representative examples
Machine learning	Random forest, LSTM[a], CNN[b], gradient boosting	20	Vaskovsky and Chvanova et al [22], Shamanna et al [15], Colmegna et al [23], Batagov et al [12]
Statistical or probabilistic methods	Logistic regression, Bayesian inference, survival analysis	9	Shamanna et al [15], Leszczełowska et al [31], Zhu et al [33], Vaskovsky et al [34]
Physiological modeling	ODEs[c], mechanistic models, compartmental models	8	Cappon et al [9], Ahmadasas et al [10], Young et al [24], Wang et al [26]
Control theory	MPC[d], optimal control, PID[e] controllers	3	Ahmadasas et al [10], Wang et al [26], Zavitsanou et al [27]
Reinforcement learning	DQN[f], Soft Actor–Critic	2	Sarani Rad et al [3], Chen et al [30]
Control, estimation, or signal processing	Kalman filtering, signal estimation, signal processing algorithms	3	Ahmadasas et al [10], Zavitsanou et al [27], Vaskovsky et al [34]
Optimization or model calibration	Parameter estimation, parameter fitting algorithms	2	Ahmadasas et al [10], Thamotharan et al [25]
Rule-based systems	Dynamic risk thresholding, equation-based bolus calculation, rule-based reasoning	2	Pellizzari et al [29], Zhu et al [33]
Simulation-based modeling	Simulation training, Euler's method	2	Mishra et al [28], Chen et al [30]
Natural language processing	GPT[g]-based natural language generation	1	Cappon et al [36]
System dynamics	Causal loop diagrams, feedback modeling	1	Mishra et al [28]

[a]LSTM: long short-term memory.
[b]CNN: convolutional neural network.
[c]ODE: ordinary differential equation.
[d]MPC: model predictive control ___.
[e]PID: proportional-integral-derivative__
[f]DQN: deep Q-network___.
[g]GPT: generative pre-trained transformer.

Key findings included:

- ML was the most common approach (20, 71.4% studies), used for glucose prediction, patient modeling, and feature extraction [3,7-9,11-18,22,23,27,31-35].
- Statistical and probabilistic methods appeared in 9 (32.1%) studies, often applied to regression, inference, or survival analysis [3,7,8,15,23,31,33,34,36].
- Physiological modeling was reported in 8 (28.6%) studies, leveraging ordinary differential equations, compartmental models, and mechanistic representations [9,10,24-26,29,30,36].
- Control-based approaches were less frequent, with control theory (3, 10.7% studies) [10,26,27] and reinforcement learning (2, 7.1% studies) supporting adaptive insulin delivery and personalization [3,30].

Target Conditions and Use Context (RQ4, RQ5)

Overview

This section summarizes the specific types of diabetes addressed in DT studies (RQ4) and the clinical goals these models aim to support (RQ5). Together, these questions provide insight into intended use cases and patient populations for DT applications in diabetes care.

Target Conditions (RQ4)

DT studies in diabetes addressed multiple forms of the disease, with some models applicable to more than 1 type. Most studies focused on type 1 diabetes (T1D) or type 2 diabetes (T2D), whereas fewer studies targeted gestational diabetes or diabetes-related complications. Table 5 summarizes the targeted diabetes types and disease stages addressed across the included studies, with representative examples.

XSL•FO
RenderX

Table . Targeted diabetes types and disease stages addressed in included studies with representative examples (research question 4).

Diabetes type	Studies, n	Representative examples
Type 2 diabetes	14	Colmegna et al [23], Mishra et al [28], Villa-Tamayo et al [32], Shamanna et al [18]
Type 1 diabetes	13	Thamotharan et al [25], Wang et al [26], Zavitsanou et al [27], Batagov et al [12]
Diabetic retinopathy (secondary to diabetes)	1	Chahal et al [35]
Gestational diabetes	1	Leszczełowska et al [31]

Key findings included:

- T2D was the most frequent target (14, 50% studies), with models supporting therapy optimization, metabolic simulation, and lifestyle interventions [3,7-9,13-18,23,28,32,34].
- T1D was addressed in 13 (46.4%) studies, primarily through closed-loop systems, real-time insulin delivery, and glucose control simulations [3,10-12,22,24-27,29,30,33,36].
- Diabetic complications were rarely considered, with 1 (3.6%) study focused on diabetic retinopathy [35].
- Gestational diabetes was examined in 1 (3.6%) study, reflecting limited application to pregnancy-related diabetes [31].

Clinical Goals (RQ5)

DT applications in diabetes addressed a broad range of clinical objectives, spanning real-time monitoring, safety, decision support, and long-term disease management. These goals were classified into primary categories reflecting their roles in clinical care. Table 6 summarizes the clinical goals of diabetes DTs, including their functions and representative examples.

Table . Clinical applications of digital twin models for diabetes care with representative examples from included studies (research question 5). Percentages may exceed 100% because individual studies could be coded into more than one category.

Clinical goal category	Key characteristics	Studies, n	Representative examples
Therapeutic control or intervention	Insulin dosing, glycemic variability management, closed-loop control	17	Colmegna et al [23], Wang et al [26], Zavitsanou et al [27], Shamanna et al [18]
Decision support or treatment planning	Dietary recommendation, therapy optimization, clinician support	10	Colmegna et al [23], Thamotharan [25], Mishra et al [28], Cappon et al [36]
Safety or alerting system	Hypoglycemia alerts, early glycemic warnings, safety enhancement	10	Young et al [24], Pellizzari et al [29], Chen et al [30], Joshi et al [11]
Disease prediction or forecasting	Glucose forecasting, disease progression prediction, GDM[a] risk	9	Sarani Rad et al [3], Shammana et al [8], Zhang et al [7], Joshi et al [11]
Disease management or remission	HbA_{1c}[b] reduction, weight loss, medication reduction	8	Shamanna et al [8], Shamanna et al [14], Shamanna et al [16], Surian et al [13]
Monitoring or control	Glucose time-in-range, health monitoring, normoalbuminuric	6	Sarani Rad et al [3], Young et al [24], Mishra et al [28], Shamanna et al [16]
Diagnosis or screening	DR[c] detection, GDM diagnosis, complication screening	4	Zhang et al [7], Leszczełowska et al [31], Vaskovsky et al [34], Chahal et al [35]
Risk assessment	Maternal risk, risk stratification	3	Leszczełowska et al [31], Villa-Tamayo et al [32], Vaskovsky et al [34]

[a]GDM: gestational diabetes mellitus.

[b]HbA_{1c}: hemoglobin A_{1c}.

[c]DR: diabetic retinopathy.

Key findings included:

- Therapeutic control or intervention was the most common application (17, 60.7% studies), including insulin dosing, closed-loop control, and management of glycemic variability [3,8,10,13-18,23,25-27,29,30,33,36].
- Decision support and treatment planning were reported in 10 (35.7%) studies, covering dietary recommendations, therapy optimization, and clinician-facing guidance [3,7-9,23-25,28,33,36].

XSL•FO
RenderX

- Safety and alerting systems also appeared in 10 (35.7%) studies, emphasizing hypoglycemia warnings and proactive risk alerts [10-12,24-27,29,30,34].
- Disease prediction or forecasting was described in 9 (32.1%) studies, targeting HbA_{1c} trajectories, disease progression, and gestational diabetes risk [3,7-9,11,12,22,31,32].

In Table 6, "Therapeutic control or intervention" refers to systems that actively optimize or recommend treatment actions, such as insulin dosing or therapy adjustment; "monitoring or control" refers to systems focused on tracking glycemic status or physiological trends; and "decision support or treatment planning" refers to systems that inform clinician or patient decision-making without necessarily acting as real-time controllers.

Data Sources and Personalization Mechanisms (RQ6, RQ7)

Overview

This section summarizes the types of data used to construct or update DTs for diabetes (RQ6) and the mechanisms through which these models enable personalization or self-management (RQ7). These aspects reflect both the technical input and patient-centered application of DT systems.

Data Sources (RQ6)

DT models drew on a wide range of data sources to ensure an accurate representation of patient state and dynamics. These included lifestyle, sensor-derived, clinical, and synthetic datasets, with varying degrees of adoption across studies. Table 7 summarizes the data sources used in diabetes DT systems, with representative examples.

Table . Clinical and behavioral data sources used in digital twin systems with representative examples (research question 6). Percentages may exceed 100% because individual studies could be coded into more than one category.

Data category	Key characteristics	Studies, n	Representative examples
Lifestyle data	Physical activity, dietary intake, sleep patterns	20	Colmegna et al [23], Zavitsanou et al [27] , Shamanna et al [17], Shamanna et al [18]
Wearable devices	Heart rate, insulin delivery data, blood pressure,	19	Sarani Rad et al [3], Chen et al [30], Shamanna et al [16], Shamanna et al [17]
CGM[a]	CGM data, blood glucose measurements, glucose monitors	18	Vaskovsky and Chvanova [22], Colmegna et al [23], Shamanna et al [17], Shamanna et al [18]
Electronic health records	Clinical history, laboratory results, patient demographics	12	Shamanna et al [15], Villa-Tamayo et al [32], Shamanna et al [17], Shamanna et al [18]
Simulated and public datasets	PIMA[b] dataset, UVa/Padova simulator, synthetic NHANES[c] data	6	Wang et al [26], Zavitsanou et al [27], Mishra et al [28], Chahal et al [35]
Physiological parameters	Body weight, personal characteristics, physiological metrics	5	Ahmadasas et al [10], Thamotharan et al [25], Zavitsanou et al [27], Pellizzari et al [29]
Patient-reported outcomes	Mobile health logs, self-monitoring, patient-generated input	3	Sarani Rad et al [3], Zhang et al [7], Pellizzari et al [29]
Genomic data	Metabolomics, proteomics	1	Zhang et al [7]
Imaging data	Fundus images, Optos scans, Gaussian-filtered visuals	1	Chahal et al [35]

[a]CGM: continuous glucose monitoring.

[b]PIMA: Pima Indians Diabetes Dataset.

[c]NHANES: National Health and Nutrition Examination Survey.

Key findings included:

- Lifestyle data were the most widely used input (20, 71.4% studies), covering physical activity, dietary intake, sleep, and behavioral logs [8-18,22-27,30,33,36].
- Wearable devices were incorporated in 19 (67.9%) studies, capturing heart rate, insulin delivery, and blood pressure [3,10-12,15-18,22-27,29,30,33,34,36].
- CGM appeared in 18 (64.3%) studies, enabling real-time tracking, control feedback, and risk forecasting [8,10-18,22-25,29,30,33,36].
- EHRs were used in 12 (42.9%) studies, providing longitudinal medical history, laboratory results, and medication data [3,7-9,13-18,32,33].
- Synthetic and public datasets were used in 6 (21.4%) studies, often for simulation or benchmarking, such as the UVa/Padova simulator or National Health and Nutrition Examination Survey (NHANES) data [26-28,31,32,35].

XSL•FO
RenderX

Personalization Mechanisms (RQ7)

Most DT systems aimed to enable personalized care through individualized feedback, adaptive modeling, or real-time decision support. Personalization strategies varied in scope, ranging from lifestyle guidance to therapy optimization and digital coaching. Table 8 summarizes the personalization features and tailoring strategies used in diabetes DT systems, with representative examples.

Table . Personalized features and patient-specific tailoring strategies in digital twin systems with representative examples (research question 7). Percentages may exceed 100% because individual studies could be coded into more than one category.

Personalization mechanism category	Key characteristics	Studies, n	Representative studies
Personalized lifestyle recommendations	Nutrition guidance, individualized meal or activity plans, lifestyle support	11	Cappon et al [9], Young et al [24], Shamanna et al [17], Shamanna et al [18]
Real-time or adaptive personalization	Dynamic feedback, CGM[a]-based tuning, adaptive intervention planning	11	Chen et al [30], Leszczełowska et al [31], Vaskovsky et al [34], Chahal et al [35]
Personalized insulin or therapy optimization	Personalized virtual patients, Replay-BG, health scenario simulation	10	Shamanna et al [15], Ahmadasas et al [10], Zhu et al [33], Cappon et al [36]
Self-management tools or patient interface	App feedback, color-coded food systems, personalized tracking tools	8	Sarani Rad et al [3], Shamanna et al [16], Shamanna et al [17], Surian et al [13]
Individualized simulation models	Personalized virtual patients, Replay-BG, health scenario simulation	6	Sarani Rad et al [3], Zavitsanou et al [27], Pellizzari et al [29], Chen et al [30]
Behavior-driven personalization	AI-guided nudges, digital coaching, human support	4	Shamanna et al [14], Colmegna et al [23], Shamanna et al [16], Surian et al [13]
Safety or alerting system	Tailored alerts, risk-specific notifications	1	Vaskovsky et al [34]

[a]CGM: continuous glucose monitoring.

Key findings:

- Personalized lifestyle recommendations were the most frequent approach (11, 39.3% studies), providing tailored nutrition, activity, and daily routine guidance [3,8,9,13-18,23,24].
- Real-time or adaptive personalization was also reported in 11 (39.3% studies), offering interventions dynamically responsive to CGM and sensor feedback [8,10,13,15,17,27,30,31,33-35].
- Personalized insulin or therapy optimization appeared in 10 (35.7% studies), focusing on precision dosing, medication planning, and adaptive therapy [3,9,10,15,18,23-25,33,36].
- Individualized simulation models were described in 6 (21.4%) studies, enabling patient-specific scenario testing and comparative evaluation [3,27,29-31,36].

Intelligence and Adaptability (RQ8)

Overview

This section explores how DT systems in diabetes manage uncertainty, real-time data updates, and interpretability. These features are central to ensuring the trustworthiness, safety, and clinical relevance of DT models in dynamic health care settings.

Handling Uncertainty, Adaptation, and Interpretability (RQ8)

Based on the 28 included studies, 5 main categories of strategies were identified. Table 9 summarizes the strategies used for handling uncertainty, dynamic adaptation, and interpretability in diabetes DT systems.

XSL•FO
RenderX

Table . Reported strategies for managing uncertainty, real-time dynamics, and interpretability in diabetes digital twin models with representative examples (research question 8). Percentages may exceed 100% because individual studies could be coded into more than one category.

Strategy category	Key characteristics	Studies, n	Representative examples
Adaptive learning	Feedback loop tuning, model retraining, dynamic personalization	18	Shamanna et al [15], Thamotharan et al [25], Mishra et al [28], Vaskovsky et al [34]
Explainable AI[a]	Feature importance, knowledge graphs, visual interpretability	16	Vaskovsky and Chvanova [22], Colmegna et al [23], Wang et al [26], Chahal et al [35]
Real-time synchronization	Real-time CGM[b] updates, Kalman filtering, continuous data sync	15	Colmegna et al [23], Ahmadasas et al [10], Wang et al [26], Chahal et al [35]
Confidence scoring	Cross-validation, confidence intervals, robustness testing	12	Vaskovsky and Chvanova [22], Zavitsanou et al [27], Mishra et al [28], Zhang et al [7]
Human-in-the-loop	Physician monitoring, manual oversight, feedback mechanisms	3	Shamanna et al [14], Mishra et al [28], Shamanna et al [18]

[a]AI: artificial intelligence.

[b]CGM: continuous glucose monitoring.

Key findings included:

- Adaptive learning was the most common capability (18, 64.3% studies), enabling dynamic personalization through feedback loop tuning, model retraining, and continuous parameter updates [3,8-10,13-18,23,25-28,30,33,34].
- Explainable AI appeared in 16 (57.1%) studies, using methods such as feature importance analysis, visual interpretability, and knowledge graphs to improve transparency [3,7,10-12,15,22,23,25,26,29,32-36].
- Real-time synchronization was reported in 15 (53.6%) studies, supporting continuous data integration from CGM and other sensors via Kalman filtering and real-time updates [3,8,10-14,16,18,23,26,27,33-35].
- Confidence scoring approaches were applied in 12 (42.9%) studies, employing cross-validation, CIs, and robustness testing to quantify uncertainty [7-9,11,12,15,22,27,28,30,32,36].
- Human-in-the-loop oversight was reported in 3 (10.7%) studies, providing physician monitoring or manual intervention in safety-critical contexts [14,18,28].

Evaluation and Validation (RQ9, RQ10)

Overview

This section summarizes reported outcomes from DT applications in diabetes (RQ9) and describes the methods used to validate these systems (RQ10). Together, these questions address the effectiveness and credibility of DT models in clinical and experimental contexts.

Reported Outcomes (RQ9)

Across the 28 included studies [3,7-18,22-36], reported outcomes varied widely depending on the DT system's clinical target and implementation maturity. Outcomes were grouped into major categories reflecting both clinical and system-level effects. Table 10 summarizes the clinical outcomes of DTs for diabetes.

XSL•FO
RenderX

Table . Clinical outcomes reported in digital twin research for diabetes, categorized by outcome type with representative examples (research question 9). Percentages may exceed 100% because individual studies could be coded into more than one category.

Outcome category	Key characteristics	Studies, n	Representative examples
Improved HbA_{1c}[a] or glycemic control	Increased time in range, HbA_{1c} reduction, improved control	17	Cappon et al [9], Thamotharan et al [25], Wang et al [26], Chen et al [30]
Other clinical benefits	Retinopathy or nephropathy improvement, cardiovascular risk reduction	11	Shamanna et al [8], Colmegna et al [23], Leszczełowska et al [31], Villa-Tamayo et al [32]
Improved prediction accuracy	Accurate glucose or GDM[b] prediction, low RMSE[c] or MAE[d]	9	Vaskovsky and Chvanova [22], Zavitsanou et al [27], Leszczełowska et al [31], Chahal et al [35]
Medication use reduction	Reduced or discontinued medication use	6	Shamanna et al [14], Shamanna et al [17], Shamanna et al [18], Surian et al [13]
Weight or metabolic outcomes	Weight loss, improved insulin resistance, BMI reduction	5	Shamanna et al [8], Shamanna et al [14], Shamanna et al [17], Surian et al [13]
Hypo- or hyperglycemia reduction	Fewer glycemic events, improved variability	5	Thamotharan et al [25], Zavitsanou et al [27], Chen et al [30], Zhu et al [33]
T2D[e] remission or reversal	Diabetes remission or reversal	3	Shamanna et al [8]. Shamanna et al [15], Surian et al [13]
Improved detection or screening	Higher detection rates, classification accuracy	2	Mishra et al [28], Vaskovsky et al [34]
Blood pressure outcomes	Hypertension remission, reduced SBP[f]/DBP[g]	2	Shamanna et al [8], Shamanna et al [16]
Early detection or decision support	Improved early intervention	1	Vaskovsky et al [34]
Enhanced patient engagement	Improved patient comprehension and engagement	1	Sarani Rad et al [3]
Patient or clinician satisfaction	High clinician satisfaction	1	Zhu et al [33]
Personalized therapy optimization	Enhanced insulin dosing precision	1	Ahmadasas et al [10]

[a]HbA_{1c}: hemoglobin A_{1c}.
[b]GDM: gestational diabetes mellitus.
[c]RMSE: root mean square error.
[d]MAE: mean absolute error.
[e]T2D: type 2 diabetes.
[f]SBP: systolic blood pressure.
[g]DBP: diastolic blood pressure.

Key findings:

- Improved HbA_{1c} or glycemic control was the most frequently reported outcome (17, 60.7% studies), showing HbA_{1c} reduction, increased TIR, and reduced variability [3,8-10,13-15,17,18,23-27,29,30,33].
- Other clinical benefits were described in 11 (39.3%) studies, including retinopathy or nephropathy improvement and cardiovascular risk reduction [8,15-17,22,23,27,31,32,34,35].
- Improved prediction accuracy was reported in 9 (32.1%) studies, with accurate glucose or gestational diabetes mellitus prediction and low root-mean-square error (RMSE) and mean absolute error (MAE) [3,7,9,11,12,22,31,32,35].
- Less frequently, outcomes included medication use reduction, weight or metabolic improvements, hypo- or hyperglycemia reduction, and other patient-centered measures.

Reported quantitative outcomes suggest that some DT applications were associated with clinically meaningful improvements, although results varied by study design and use case. In 1 retrospective T2D cohort, HbA_{1c} decreased from 8.8% to 6.9% after 90 days, corresponding to a 1.9 percentage-point reduction, together with a 56.9% reduction in homeostatic model assessment of insulin resistance, a 6.1% decrease in body weight, and 89.1% (57/64) of participants achieving time in range (70 - 180 mg/dL) ≥70% after the intervention [18]. In a DT-based exercise decision support system for T1D, mean time in range improved from 80.2% to 92.3% for aerobic exercise

XSL•FO
RenderX

and from 72.3% to 87.3% for resistance exercise, while time spent in low glucose decreased from 15.1% to 5.1% and from 18.2% to 6.6%, respectively [24]. A mechanistic personalized nutrition model in prediabetes predicted individual body weight and HbA_{1c} trajectories with mean prediction errors of 0.7 kg and 0.08 percentage points in the training dataset, and approximately 1.1% and 1.4% percentage errors, respectively, in the test dataset [30]. Some prediction-focused systems also reported strong performance metrics, including RMSE 24.96 mg/dL, MAE 17.21 mg/dL, and area under the receiver operating characteristic curve >0.85 for postprandial glucose prediction, as well as area under the curve (AUC) of 0.80 - 0.82 for chronic kidney disease identification and AUC 0.86 for 3-year chronic kidney disease prediction in T2D cohorts [8,13]. In maternal-risk applications, 1 DT system reported 83.5% accuracy for maternal health risk assessment and 97.2% precision for gestational diabetes prediction [31].

Validation Methods (RQ10)

Validation approaches were grouped into 5 broad categories, reflecting how DT systems were evaluated for performance, safety, and generalizability. Table 11 summarizes the validation methods used in diabetes DT systems.

Table . Validation methods used in diabetes digital twin studies with representative examples (research question 10). Percentages may exceed 100% because individual studies could be coded into more than one category.

Validation method category	Key characteristics	Studies, n	Representative examples
Quantitative evaluation	Accuracy metrics (eg, RMSE[a] and AUC[b]), statistical tests, cross-validation	21	Vaskovsky and Chvanova [22], Mishra et al [28], Leszczełowska et al [31], Vaskovsky et al [34]
Retrospective validation	Cross-validation, train or test split, retrospective data analysis	10	Thamotharan et al [25], Joshi et al [11], Villa-Tamayo et al [32], Batagov et al [12]
Simulation testing	ReplayBG or UVa/Padova simulation, virtual cohort evaluation	9	Young et al [24], Wang et al [26], Pellizzari et al [29], Chen et al [30]
Clinical trials	Randomized controlled trial, pilot study, prospective design	4	Shammana et al [8], Shamanna et al [16], Zhu et al [33], Cappon et al [36]
Real-world validation	Clinical evaluation, patient outcomes, CGM[c] tracking	4	Shamanna et al [8], Colmegna et al [23], Zhu et al [33], Surian et al [13]
Expert review	Case study evaluation, user feedback	2	Shamanna et al [23], Zhu et al [33]

[a]RMSE: root-mean-square error.

[b]AUC: area under the curve.

[c]CGM: continuous glucose monitoring.

Key findings included:

- Quantitative evaluation was the most common approach (21, 75% studies), typically using accuracy metrics (eg, RMSE, MAE, and AUC) and cross-validation methods to assess performance [7-9,11-18,22,24,27-32,34,35].
- Retrospective validation was applied in 10 (35.7%) studies, using historical datasets (eg, EHRs and CGM logs) for training or testing and retrospective analysis [8,9,11,12,22,25,31,32,34,35].
- Simulation testing was reported in 9 (32.1%) studies, often leveraging tools, such as the UVa/PADOVA simulator or ReplayBG, to validate insulin control and metabolic models [3,8,10,24-27,29,30].
- Clinical and real-world evaluation was limited, with clinical evaluation reported in 4 studies (14.3%) [8,16,33,36] and real-world evaluation reported in 4 (14.3%) studies [8,13,23,33], including small-scale pilots, randomized controlled trials, or deployment in real patient settings with CGM tracking.
- Expert review was rarely used, reported in 2 (7.1%) studies, based on clinician or user feedback or case study evaluation [23,33].

Implementation and Governance (RQ11, RQ12)

Overview

This section describes how ethical, legal, and practical considerations are addressed in the implementation of DT systems for diabetes. It summarizes reported privacy and regulatory strategies (RQ11) and examines technical and workflow-related barriers to deployment (RQ12). Together, these questions assess readiness for safe, responsible, and scalable clinical integration.

Privacy, Ethical, and Regulatory Considerations (RQ11)

DT systems introduce complex ethical and legal considerations due to their reliance on sensitive health data and AI-driven decision-making. Among the 28 studies [3,7-18,22-36], 4 high-level categories were identified—data privacy, consent and transparency, accountability, and bias or fairness. Table 12 summarizes the strategies used for handling privacy, ethical, and regulatory issues in diabetes DT systems.

XSL•FO
RenderX

Table . Ethical, privacy, and regulatory considerations in diabetes digital twins with representative examples (research question 11). Percentages may exceed 100% because individual studies could be coded into more than one category.

Ethics or privacy category	Key characteristics	Studies, n	Representative examples
Data privacy	Data anonymization, encryption, GDPR[a] or HIPAA[b] compliance	8	Mishra et al [28], Zhu et al [33], Vaskovsky et al [34], Chahal et al [35]
Accountability	Audit trails, regulatory compliance, and interoperability	6	Cappon et al [9], Zhu et al [33], Vaskovsky et al [34], Chahal et al [35]
Consent and transparency	Data ownership, ethics approval obtained, informed consent, patient consent, permission-based data storage	6	Zhu et al [33], Vaskovsky et al [34], Chahal et al [35], Cappon et al [36]
Bias and fairness	Identification of bias potential	1	Sarani Rad et al [3]

[a]GDPR: General Data Protection Regulation.

[b]HIPAA: Health Insurance Portability and Accountability Act.

Key findings included:

- Data privacy was the most frequently discussed (8, 28.6% studies), typically through anonymization, encryption, and compliance with HIPAA (Health Insurance Portability and Accountability Act) or GDPR (General Data Protection Regulation) [3,8,9,28,33-36].
- Accountability appeared in 6 (21.4%) studies, including the use of audit trails, traceability, and regulatory compliance mechanisms [7-9,33-35].
- Consent and transparency were also reported in 6 (21.4%) studies, covering informed consent procedures, institutional review board approvals, and patient-facing disclosures [7,8,33-36].
- Bias and fairness were noted in only 1 (3.6%) study, reflecting a critical underexplored gap in addressing algorithmic inequity [3].

Implementation Barriers and Enablers (RQ12)

Although many DT systems demonstrated technical feasibility, real-world implementation remains constrained by several recurring challenges. These were grouped into 4 main categories—data quality and availability, model limitations, validation limitations, and workflow or interoperability barriers. Table 13 summarizes the implementation barriers that exist in diabetes DT systems.

Table . Implementation barriers and enablers in diabetes digital twin systems, with representative examples (research question 12). Percentages may exceed 100% because individual studies could be coded into more than one category.

Implementation barriers category	Key characteristics	Studies, n	Representative examples
Validation limitation	Lack of randomization, short follow-up, and personalization gaps	16	Wang et al [26], Zavitsanou et al [27], Villa-Tamayo [32], Shamanna et al [18]
Data quality or availability	Burden of data collection, missing variables, limited real-world data, and synthetic datasets	14	Valovsky and Chvanova [22], Wang et al [26], Shamanna et al [17], Shamanna et al [18]
Model limitations	Simplified physiology, tuning complexity, and selection bias	11	Ahmadasas et al [10], Wang et al [26], Leszczełowska et al [31], Villa-Tamayo et al [32]
Workflow and interoperability	Clinical workflow alignment, data format compatibility issues, data integration challenges, data integration from multiple sources, integration with existing devices, and interoperability challenges	8	Cappon et al [9], Colmegna et al [23], Mishra et al [28], Zhu et al [33]

Key findings included:

- Validation limitations were the most common barrier (16, 57.1% studies), reflecting reliance on synthetic datasets, short follow-up durations, and lack of external clinical evaluation [8,10-18,25-27,30,32,33].
- Data quality and availability issues were reported in 14 (50%) studies, including missing data, unreliable sensors, and burdensome data collection procedures [7,11,15,17,18,22,25,26,29-32,35,36].
- Model limitations were described in 11 (39.3%) studies, such as limited personalization, oversimplified physiological modeling, or small training datasets [3,7,10,13,14,16,26,27,29,31,32].
- Workflow and interoperability barriers appeared in 8 (28.6%) studies, emphasizing difficulties integrating DTs

XSL•FO
RenderX

into clinical workflows, EHR systems, and device ecosystems [9,10,23,28,33-36].

Research and Development Gaps (RQ13)

Although DT systems for diabetes are showing technical feasibility, multiple areas require further investigation and refinement. From the 28 reviewed studies [3,7-18,22-36], seven major gap categories that emerged were (1) limited scope of application, (2) integration challenges, (3) lack of longitudinal data, (4) data quality and availability, (5) methodological limitations, (6) need for clinical validation, and (7) scalability or usability concerns. Table 14 summarizes the reported research and development gaps in diabetes DT systems.

Table . Reported research gaps and future development needs in diabetes digital twin literature with representative examples (research question 13). Percentages may exceed 100% because individual studies could be coded into more than one category.

Gap category	Key characteristics	Studies, n	Representative examples
Need for clinical validation	Larger clinical trials and subgroup and demographic validation	15	Shamanna et al [15], Thamotharan et al [25], Shamanna et al [18], Zhu et al [33]
Limited scope of application	Broader populations, diverse settings, and multimorbidity expansion	14	Sarani Rad et al [3], Shamanna et al [15], Cappon et al [9], Thamotharan et al [25], Zhang et al [7]
Integration challenges	Integration with EHRs[a], real-time systems, and closed-loop models	11	Thamotharan et al [25], Joshi et al [11], Zhu et al [33], Batagov et al [12]
Usability and real-world adoption	Personalization for MDI[b] users, real-world evaluation, and broader adoption	11	Ahmadasas et al [10], Wang et al [26], Vaskovsky et al [34], Chahal et al [35]
Lack of longitudinal data	Long-term outcome tracking, sustainability, and effectiveness studies	8	Shamanna et al [14], Shamanna et al [15], Cappon et al [9], Surian et al [13]
Data quality and availability	Dependence on wearable devices and data quality, expansion to broader population data, expansion to larger datasets, limitations in meal tracking and calibration, and need for denser time-series data	6	Vaskovsky and Chvanova [22], Wang et al [26], Mishra et al [28], Villa-Tamayo [32]
Methodological limitations	Standardized protocols, adaptive learning, and causal reasoning	5	Sarani Rad et al [3], Vaskovsky and Chvanova [22], Colmegna et al [23], Pellizzari et al [29]
Scalability challenges	Deployment in low-resource settings and real-world scalability	3	Leszczełowska et al [31], Zhu et al [33], Chahal et al [35]

[a]EHR: electronic health record.

[b]MDI: multiple daily injection.

Key findings included:

- Need for clinical validation was the most frequently cited gap (15, 53.6% studies), reflecting the lack of randomized trials, subgroup evaluations, and real-world testing [3,9,13-16,18,23,25,27,29,31,33,34,36].
- Limited scope of application was reported in 14 (50%) studies, with DTs often targeting narrow use cases and failing to generalize across diverse populations or multimorbidity contexts [3,7-10,13,15-17,25-28,31].
- Integration challenges were noted in 11 (39.3%) studies, underscoring difficulties with EHR interoperability, real-time deployment, and multidevice environments [7,11,12,22,25,27,30,32-35].
- Usability and real-world adoption also appeared in 11 (39.3%) studies, pointing to the need for personalization, support for multiple daily injection users, and strategies for broader adoption in routine care [8,10,12,16,24,26,30,32,34-36].

Discussion

This PRISMA-ScR–compliant scoping review maps the current state of DT systems in diabetes, addressing 13 structured research questions across 7 thematic domains.

System Design and Modeling Foundations (RQ1, RQ2, RQ3)

DT systems for diabetes use a wide range of modeling techniques, most commonly ML (eg, long short-term memory, gradient boosting, and reinforcement learning) and physiological simulation. Simulation engines and predictive ML modules were often integrated into layered architectures that also included personalization modules, decision support, and user-facing dashboards. Statistical and probabilistic methods (eg, regression and Bayesian inference) were also used in several studies, although less prominently. Few systems incorporated mechanistic control theory or signal-processing models. The inclusion of key components, such as simulation engines,

XSL•FO
RenderX

control-feedback modules, and data integration pipelines, reflects a growing maturity in system design.

Target Conditions and Use Context (RQ4 and RQ5)

Most DTs targeted T1D or T2D, with limited applications in gestational diabetes or diabetes-related complications, such as retinopathy. Primary clinical goals included glycemic prediction, insulin-dose optimization, lifestyle guidance, and therapeutic planning. Several systems also addressed the diagnosis of complications or risk stratification for comorbidities. The breadth of clinical use cases suggests that DTs are evolving from simple simulators into multifunctional clinical-support tools.

Data Sources and Personalization Mechanisms (RQ6 and RQ7)

Lifestyle data, wearable devices, and CGM were the dominant inputs, with hybrid combinations being common. EHRs and synthetic datasets were also widely used to provide historical or simulated information. Personalization was achieved through mechanisms such as real-time adaptation, individual model tuning, behavior-driven feedback (eg, nudges), and insulin titration. However, persistent challenges remain in data quality, sensor integration, and dataset heterogeneity.

Intelligence and Adaptability (RQ8)

Managing uncertainty and real-time updates is crucial for clinical reliability. Studies implemented adaptive learning, feedback loops, and explainable-AI methods (eg, attention mechanisms and knowledge graphs) to improve transparency and adaptability. Real-time CGM synchronization and, in some cases, human-in-the-loop oversight were used to enhance model responsiveness and safety.

Evaluation and Validation (RQ9 and RQ10)

Quantitative validation (eg, RMSE and AUC) was common, but real-world clinical trials were rare. Most studies validated systems via retrospective datasets or simulations. Reported clinical outcomes included improved TIR, fewer hypoglycemic events, and, in some cases, T2D remission. However, evidence on long-term effectiveness, generalizability, and cost-effectiveness remains limited.

A notable finding across the included studies is the mismatch between technical sophistication and clinical maturity. Although many DT systems incorporated adaptive learning, individualized simulation, and multimodal data integration, most were evaluated using retrospective datasets or in silico simulations rather than prospective clinical deployment. This likely reflects the high implementation burden of DTs in diabetes, including the need for reliable real-time data streams, safety safeguards, interoperability with devices and clinical systems, and acceptable workflow integration. It also reflects the regulatory complexity of systems that may influence insulin dosing or therapeutic decision-making.

Implementation and Governance (RQ11, RQ12)

Privacy and ethical considerations were addressed inconsistently, often limited to brief compliance mentions (eg, GDPR and HIPAA). A smaller subset of studies explicitly discussed accountability (eg, audit trails and governance mechanisms) or algorithmic bias and fairness, highlighting underexplored areas of governance. Implementation enablers included real-time feedback and sensor integration, whereas barriers included poor data quality, system complexity, lack of clinical workflow alignment, and limited scalability.

Another important finding is the limited attention to algorithmic bias and fairness. Despite the increasing use of AI-driven modeling and decision-support approaches, only a small subset of studies explicitly discussed bias, representativeness, or equity-related concerns. This suggests that the field is still focused primarily on technical feasibility and predictive performance rather than equitable deployment across diverse patient populations.

Research and Development Gaps (RQ13)

Key gaps include limited clinical validation, insufficient longitudinal data, a lack of standardized model architectures, and limited generalizability to diverse populations. Many studies emphasized the need for integration with EHRs, real-world testing, and regulatory alignment. Addressing these gaps will be essential to enable scalable, equitable, and clinically robust DT systems for diabetes management.

Summary and Implications

This review offers a panoramic view of the evolving DT landscape in diabetes. While notable technical advances are evident—particularly in data integration and personalization—the field remains formative, with substantial work needed in clinical validation, ethical governance, and system interoperability. Future research should emphasize not only algorithmic sophistication but also real-world applicability, safety, and equity to support the scalable and responsible deployment of DTs in diabetes care.

Taken together, the literature suggests that DT research in diabetes is progressing from conceptual and simulation-based work toward more clinically relevant systems, but the field remains early in real-world maturity. Future studies should prioritize prospective validation, broader demographic and clinical representation, transparent reporting, interoperability with routine care systems, and governance frameworks that address privacy, accountability, and fairness.

Limitations

This scoping review has several limitations. First, only English-language studies with accessible full text were included, and gray literature was excluded, which may have led to the omission of some relevant studies. Second, formal risk-of-bias and certainty-of-evidence assessments were not performed because the aim was to map a heterogeneous body of literature rather than evaluate intervention effects. Third, the included studies differed substantially in design, terminology, validation methods, and outcomes, limiting direct comparison. Finally, many studies were early-phase, retrospective, or simulation-based, which limits conclusions about clinical effectiveness and real-world implementation.

XSL•FO
RenderX

Acknowledgments

Generative artificial intelligence was used to assist with language editing and manuscript drafting. All AI-assisted output was reviewed, edited, and verified by the authors, who take full responsibility for the final content of the manuscript.

Funding

This work was supported by the National Science Foundation under grants OIA-2218046 and OIA-2611071.

Data Availability

All data analyzed in this scoping review were charted from publicly available publications. The search strategy, screening criteria, extracted study characteristics, and supplementary review materials are provided in the manuscript and its supplementary files. No primary participant-level dataset was generated for this study.

Authors' Contributions

FSR, EB, and MJ contributed to conceptualization. FSR, KK, and MJ contributed to data curation. FSR and EB contributed to formal analysis and methodology. JL contributed to supervision. FSR, KK, and MJ wrote the original draft. JL reviewed and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File, 30 KB - diabetes_v11i1e83059_app1.doc]

Multimedia Appendix 2
Excluded full-text articles and reasons for exclusion.
[DOCX File, 17 KB - diabetes_v11i1e83059_app2.docx]

Multimedia Appendix 3
Study characteristics and data-charting form.
[XLSX File, 15 KB - diabetes_v11i1e83059_app3.xlsx]

Checklist 1
PRISMA checklist.
[PDF File, 165 KB - diabetes_v11i1e83059_app4.pdf]

References

1. Emmert-Streib F. Defining a digital twin: a data science-based unification. MAKE 2023;5(3):1036-1054. [doi: 10.3390/make5030054]
2. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care: ethical implications of an emerging engineering paradigm. Front Genet 2018;9:31. [doi: 10.3389/fgene.2018.00031] [Medline: 29487613]
3. Sarani Rad F, Hendawi R, Yang X, Li J. Personalized diabetes management with digital twins: a patient-centric knowledge graph approach. J Pers Med 2024 Mar 28;14(4):359. [doi: 10.3390/jpm14040359] [Medline: 38672986]
4. American Diabetes Association. Standards of medical care in diabetes. Diabetes Care 2005 Jan;28 Suppl 1(Suppl 1):S4-S36. [doi: 10.2337/diacare.28.suppl_1.S4] [Medline: 15618112]
5. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2021. Diabetes Care 2021 Jan;44(Suppl 1):S73-S84. [doi: 10.2337/dc21-S006] [Medline: 33298417]
6. Sun T, He X, Li Z. Digital twin in healthcare: recent updates and challenges. Digit Health 2023;9:20552076221149651. [doi: 10.1177/20552076221149651] [Medline: 36636729]
7. Zhang Y, Qin G, Aguilar B, et al. A framework towards digital twins for type 2 diabetes. Front Digit Health 2024;6:1336050. [doi: 10.3389/fdgth.2024.1336050] [Medline: 38343907]
8. Shamanna P, Joshi S, Thajudeen M, et al. Personalized nutrition in type 2 diabetes remission: application of digital twin technology for predictive glycemic control. Front Endocrinol (Lausanne) 2024;15:1485464. [doi: 10.3389/fendo.2024.1485464] [Medline: 39634180]

XSL•FO
RenderX

9. Cappon G, Vettoretti M, Sparacino G, Favero SD, Facchinetti A. ReplayBG: a digital twin-based methodology to identify a personalized model from type 1 diabetes data and simulate glucose concentrations to assess alternative therapies. IEEE Trans Biomed Eng 2023 Nov;70(11):3227-3238. [doi: 10.1109/TBME.2023.3286856] [Medline: 37368794]
10. Ahmadasas M, Rashid MM, Siket M, Abdel-Latif MM, Shahidehpour A, Cinar A. Personalized artificial pancreas for glucose regulation in people with diabetes. IFAC-PapersOnLine 2024;58(30):55-60. [doi: 10.1016/j.ifacol.2025.01.156]
11. Joshi S, Shamanna P, Dharmalingam M, et al. Digital twin-enabled personalized nutrition improves metabolic dysfunction-associated fatty liver disease in type 2 diabetes: results of a 1-year randomized controlled study. Endocr Pract 2023 Dec;29(12):960-970. [doi: 10.1016/j.eprac.2023.08.016] [Medline: 37778441]
12. Batagov A, Dalan R, Wu A, Lai W, Tan CS, Eisenhaber F. Generalized metabolic flux analysis framework provides mechanism-based predictions of ophthalmic complications in type 2 diabetes patients. Health Inf Sci Syst 2023 Dec;11(1):18. [doi: 10.1007/s13755-023-00218-x] [Medline: 37008895]
13. Surian NU, Batagov A, Wu A, et al. A digital twin model incorporating generalized metabolic fluxes to identify and predict chronic kidney disease in type 2 diabetes mellitus. NPJ Digit Med 2024 May 24;7(1):140. [doi: 10.1038/s41746-024-01108-6] [Medline: 38789510]
14. Shamanna P, Erukulapati RS, Shukla A, et al. One-year outcomes of a digital twin intervention for type 2 diabetes: a retrospective real-world study. Sci Rep 2024 Oct 26;14(1):25478. [doi: 10.1038/s41598-024-76584-7] [Medline: 39461977]
15. Shamanna P, Joshi S, Shah L, et al. Type 2 diabetes reversal with digital twin technology-enabled precision nutrition and staging of reversal: a retrospective cohort study. Clin Diabetes Endocrinol 2021 Nov 15;7(1):21. [doi: 10.1186/s40842-021-00134-7] [Medline: 34776010]
16. Shamanna P, Joshi S, Dharmalingam M, et al. Digital twin in managing hypertension among people with type 2 diabetes: 1-year randomized controlled trial. JACC Adv 2024 Sep;3(9):101172. [doi: 10.1016/j.jacadv.2024.101172] [Medline: 39372467]
17. Shamanna P, Dharmalingam M, Sahay R, et al. Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the Digital Twin Precision Treatment Program. Sci Rep 2021 Jul 21;11(1):14892. [doi: 10.1038/s41598-021-94339-6] [Medline: 34290310]
18. Shamanna P, Saboo B, Damodharan S, et al. Reducing HbA1c in type 2 diabetes using digital twin technology-enabled precision nutrition: a retrospective analysis. Diabetes Ther 2020 Nov;11(11):2703-2714. [doi: 10.1007/s13300-020-00931-w] [Medline: 32975712]
19. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
20. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
21. Jafarpour M, Bitaraf E, Moeini A, Nahvijou A. Triple A (AAA): a tool to analyze scientific literature metadata with complex network parameters. Presented at: 2023 9th International Conference on Web Research (ICWR); May 3-4, 2023; Tehran, Iran, Islamic Republic of p. 342-345. [doi: 10.1109/ICWR57742.2023.10139229]
22. Vaskovsky AM, Chvanova MS. Designing the neural network for personalization of food products for persons with genetic president of diabetic sugar. Presented at: 2019 3rd School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR); Sep 9-11, 2019; Innopolis, Russia p. 175-177. [doi: 10.1109/DCNAIR.2019.8875559]
23. Colmegna P, Wang K, Garcia-Tirado J, Breton MD. Mapping data to virtual patients in type 1 diabetes. Control Eng Pract 2020 Oct;103:104605. [doi: 10.1016/j.conengprac.2020.104605]
24. Young G, Dodier R, Youssef JE, et al. Design and in silico evaluation of an exercise decision support system using digital twin models. J Diabetes Sci Technol 2024 Mar;18(2):324-334. [doi: 10.1177/19322968231223217] [Medline: 38390855]
25. Thamotharan P, Srinivasan S, Kesavadev J, et al. Human digital twin for personalized elderly type 2 diabetes management. J Clin Med 2023 Mar 7;12(6):2094. [doi: 10.3390/jcm12062094] [Medline: 36983097]
26. Wang Q, Molenaar P, Harsh S, et al. Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: an extended Kalman Filter approach. J Diabetes Sci Technol 2014 Mar;8(2):331-345. [doi: 10.1177/1932296814524080] [Medline: 24876585]
27. Zavitsanou S, Mantalaris A, Georgiadis MC, Pistikopoulos EN. In silico closed-loop control validation studies for optimal insulin delivery in type 1 diabetes. IEEE Trans Biomed Eng 2015 Oct;62(10):2369-2378. [doi: 10.1109/TBME.2015.2427991] [Medline: 25935026]
28. Mishra V, Koul S, Taylor IW. Digital twin for diabetes management using system dynamics simulation: the case of India. 2024 Presented at: International Conference on Computational Intelligence in Communications and Business Analytics; Jan 24-26, 2024; Patna, India p. 305-313. [doi: 10.1007/978-3-031-81342-9_24]
29. Pellizzari E, Prendin F, Cappon G, Sparacino G, Facchinetti A. drCORRECT: an algorithm for the preventive administration of postprandial corrective insulin boluses in type 1 diabetes management. J Diabetes Sci Technol 2025 May;19(3):711-721. [doi: 10.1177/19322968231221768] [Medline: 38158565]
30. Chen JH, Fukasawa M, Sakane N, et al. Optimization of nutritional strategies using a mechanistic computational model in prediabetes: application to the J-DOIT1 study data. PLoS ONE 2023;18(11):e0287069. [doi: 10.1371/journal.pone.0287069] [Medline: 38033033]

XSL•FO
RenderX

31. Leszczełowska P, Mazur-Milecka M, Kowalczyk N, Sobotka M. Maternal health risk assessment using digital twin application. 2024 Presented at: 2024 16th International Conference on Human System Interaction (HSI); Jul 8-11, 2024. [doi: 10.1109/HSI61632.2024.10613547]
32. Villa-Tamayo MF, Pavan J, Breton M. Analysis on the practical identifiability of the subcutaneous oral glucose minimal model. IFAC-PapersOnLine 2024;58(24):269-274. [doi: 10.1016/j.ifacol.2024.11.048]
33. Zhu T, Li K, Herrero P, Georgiou P. GluGAN: generating personalized glucose time series using generative adversarial networks. IEEE J Biomed Health Inform 2023 Oct;27(10):5122-5133. [doi: 10.1109/JBHI.2023.3271615] [Medline: 37134028]
34. Vaskovsky AM, Chvanova MS, Rebezov MB. Creation of digital twins of neural network technology of personalization of food products for diabetics. Presented at: 2020 4th Scientific School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR); Sep 7-9, 2020; Innopolis, Russia p. 251-253. [doi: 10.1109/DCNAIR50402.2020.9216776]
35. Chahal Y, Tokas R, Sharma K. Smart solution using digital twin and iot for diabetic retinopathy. 2023 Presented at: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT); Jul 6-8, 2023. [doi: 10.1109/ICCCNT56998.2023.10307620]
36. Cappon G, Pellizzari E, Cossu L, et al. System architecture of TWIN: a new digital TWIN-based clinical decision support system for type 1 diabetes management in children. 2023 Presented at: 2023 IEEE 19th International Conference on Body Sensor Networks (BSN); Oct 9-11, 2023. [doi: 10.1109/BSN58485.2023.10331272]
37. Rad FS, Jafarpour M, Bitaraf E, Khaleghdadi K, Li J. Digital twin applications in diabetes management: scoping review. Open Science Framework. URL: https://osf.io/n49xz/ [accessed 2026-06-02]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CGM: continuous glucose monitoring
DT: digital twin
EHR: electronic health record
GDPR: General Data Protection Regulation
HbA_{1c}: hemoglobin A_{1c}
HIPAA: Health Insurance Portability and Accountability Act
MAE: mean absolute error
ML: machine learning
NHANES: National Health and Nutrition Examination Survey
OSF: Open Science Framework
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RMSE: root-mean-square error
RQ: research question
T1D: type 1 diabetes
T2D: type 2 diabetes
TIR: time-in-range

Edited by I Steenstra; submitted 27.Aug.2025; peer-reviewed by M Soheili, S Nelson; revised version received 17.Apr.2026; accepted 24.Apr.2026; published 18.Jun.2026.

Please cite as:
Sarani Rad F, Jafarpour M, Bitaraf E, Khaleghdadi K, Li J
Digital Twin Applications in Diabetes Management: Scoping Review
JMIR Diabetes 2026;11:e83059
URL: https://diabetes.jmir.org/2026/1/e83059
doi:10.2196/83059

© Fatemeh Sarani Rad, Maryam Jafarpour, Ehsan Bitaraf, Katayoon Khaleghdadi, Juan Li. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 18.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction

XSL•FO
RenderX

in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Training Community Health Workers for Diabetes Management in Low- and Middle-Income Countries: Systematic Review

Anirudh Gaurang Gudlavalleti[1,2], BDS, MA; Sureshkumar Kamalakannan[2,3], BSc, MPH, PhD; Sophie Vanbelle (S)[4], MSc, PhD; Venkata Satyanarayana Murthy Gudlavalleti[2], MBBS, MSc, MD; Nicolaas C Schaper[5], MD, PhD; Giridhara R Babu[6], MBBS, MPH, PhD; Onno CP van Schayck[1], MD, PhD

[1]Department of Family Medicine, Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

[2]Department of Public Health, Pragyaan Sustainable Health Outcomes Foundation, Level 2, Kapil Kavuri Hub, Financial District, Nanakramguda, Hyderabad, India

[3]Department of Social Work, Education and Community Wellbeing, Northumbria University, Newcastle Upon Tyne, United Kingdom

[4]Methodology and Statistics, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

[5]Department of Internal Medicine, Maastricht University, Care and Public Health Research Institute (CAPHRI), Cardiovascular Research Institute Maastricht (CARIM), Maastricht, The Netherlands

[6]Department of Population Medicine, College of Medicine, QU Health, Qatar University, Doha, Qatar

Corresponding Author:
Anirudh Gaurang Gudlavalleti, BDS, MA
Department of Family Medicine, Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

Abstract

Background: Type 2 diabetes mellitus, a public health challenge, disproportionately impacts low- and middle-income countries (LMICs), accounting for 73% of global cases. Due to resource constraints, these nations have adopted task-shifting strategies using community health workers (CHWs). However, evidence on the effectiveness of training CHWs in diabetes management is limited and, at most, indirect due to the limited studies, variable training methods, and complex interventions that make it difficult to isolate training effects.

Objective: A systematic review was conducted to answer the question: Does training CHWs in type 2 diabetes improve the efficacy of diabetes screening and management at the community level in LMICs?

Methods: A total of 2 reviewers, supervised by 2 supervisors, conducted the review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. They searched databases, including PubMed, MEDLINE Ovid, Scopus, EBMR, and CINAHL, for studies published between January 2000 and April 2024, including randomized and nonrandomized controlled trials and observational studies assessing CHW training in diabetes management in LMICs. The primary outcome was the mean change in glycated hemoglobin (HbA_{1c}) percentage levels. Data were narratively synthesized for training characteristics and study outcomes, and quality was assessed using the Risk of Bias 2 and ROBINS-I tools.

Results: A total of 3387 studies were screened; 69 were eligible for full-text review, and 4 studies (3 randomized controlled trials [RCTs] and 1 observational stepped-wedge study, ~1000 patients) were included for narrative analysis. One of the 3 RCTs reported a statistically significant mean HbA_{1c} reduction of −0.24% (P=.001), but HbA_{1c} was not the primary outcome, and most patients were normoglycemic, prediabetic, or had diabetes. Other studies reported nonsignificant HbA_{1c} reductions. The risk of bias among RCTs was moderate (some concerns, 1 trial at high risk), and the observational study had a serious risk of bias. No meta-analysis was performed due to the limited number of RCTs.

Conclusions: Training CHWs in type 2 diabetes management has shown limited and, at most, indirect effects in improving glycemic control in LMIC settings. These findings are constrained by the small number of eligible studies, heterogeneity in training methodologies, and the multicomponent nature of the included interventions, with 1 trial demonstrating a statistically significant yet small reduction in HbA_{1c} (−0.24%). Our review included only 4 eligible studies with a small representation of CHWs and multicomponent interventions. Considering the limited number of eligible studies, the heterogeneity in training methodologies and study designs, and the multicomponent nature of the included interventions, the existing evidence remains inadequate to definitively conclude whether CHW training significantly improves diabetes management across LMICs. Therefore, strengthening and standardizing CHW training might be an effective strategy to enhance diabetes care in underserved settings. Future larger trials and implementation research can help maximize the impact of CHWs against the growing diabetes burden.

Trial Registration: PROSPERO CRD42022341717; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022341717

International Registered Report Identifier (IRRID): RR2-10.2196/57313

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e84508) doi:10.2196/84508

KEYWORDS

community health workers; community health workers training; diabetes management; task shifting; systematic review; low- and middle-income countries; LMICs; glycated hemoglobin; HbA1c

Introduction

Rationale

Diabetes mellitus is a growing pandemic in the global community, with 589 million people affected by it [1], while also representing one of the leading factors of mortality and morbidity globally [2]. Nearly 430 million people with diabetes, which is 73% of such patients, reside in low- and middle-income countries (LMICs) [1]. The steadily rising prevalence [3] imposes a substantial burden on health care systems due to the complexity of accessing health care in these countries [4-6]. Furthermore, individuals residing in rural areas encounter numerous barriers to accessing health care, which can be broadly categorized into 4 key areas: geographical accessibility, health care availability, financial accessibility, and the acceptability of care. Geographical accessibility refers to the ease or difficulty of reaching health care facilities, which is more pronounced in rural areas. This is due to the large distances rural patients must traverse and the limited transportation infrastructure. Health care access is limited by shortages of facilities and trained staff, especially in rural areas. Financial barriers, particularly out-of-pocket costs, prevent many people, especially those from lower socioeconomic groups, from accessing health care. Additionally, sociocultural factors like mistrust or stigma can deter patients from seeking care. These barriers disproportionately affect rural populations and those from lower socioeconomic strata [2,7-12].

Many LMICs employ local community health workers (CHWs) to improve access to health care. CHWs are usually individuals from the communities they serve, selected by those communities, and accountable to them for their work. While CHWs should receive support from the health system, they are not required to be formally integrated into its organizational structure. Additionally, their training is typically shorter than that of professional health care workers [13]. Referred to as task shifting, using CHWs for disease management is effective for infectious diseases (eg, HIV and hepatitis), especially in pregnant women, newborns, and mothers [14,15], and for noncommunicable diseases, like diabetes and cardiovascular diseases [16-25]. While most LMICs consider skilling and implementing task shifting using CHWs to improve care quality [20], they are limited by a lack of clarity on the exact nature, content, and skills of the CHWs. Evidence suggests that the quality of health care at the primary care level is associated with the limited knowledge and requisite skills among CHWs. Hence, most task-shifting initiatives also focus on imparting requisite training to CHWs [9,26-32]. Despite many CHW training programs, there is limited systematic evidence of their impact on diabetes outcomes in LMICs. While some reviews have examined CHW roles in diabetes globally, none have specifically focused on the effect of CHW training in LMIC contexts. Despite this, the evidence regarding the impact of CHW training in LMICs is still limited and mostly indirect, and no systematic review has specifically addressed this issue. Hence, this paper aims to systematically review the available evidence to answer the question: Does training CHWs in diabetes improve the efficacy of diabetes screening and management at the community level in LMICs?

Objective

We hypothesize that CHW training will improve glycemic control among patients managed by the trained CHWs. By identifying and synthesizing existing studies, this review also aims to assess whether training can be a necessary but insufficient component of a multicomponent intervention for effective CHW-led diabetes management, highlighting knowledge gaps and informing future improvements in community-based diabetes care. This review also aims to identify gaps in the reporting of CHW training.

Methods

Study Design and Registration

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1) [33,34]. The review was registered with PROSPERO (CRD42022341717) and was published previously [35].

Eligibility Criteria

Randomized controlled trials (RCTs), non-RCTs, and observational studies that assessed the training of CHWs in the management of type 2 diabetes mellitus were included. The CHWs were included based on the definition provided by the World Health Organization (WHO) [36]. The PICOS (population, intervention, comparator, outcome, and study setting) framework was used to finalize the specific inclusion criteria, which were as follows:

- Population: the population included was CHWs managing persons with type 2 diabetes from LMICs. The WHO definition of CHWs was used, which defines them as “members of the communities where they work, should be selected by the communities, should be answerable to the communities for their activities, should be supported by the health system but not necessarily be a part of its organization, and have shorter training than professional workers” [13]. Only studies in which the target population was 18 years or older were included. The baseline qualifications of the CHWs were not reported in the included studies.
- Intervention: the review included any study in which CHWs received training in diabetes management, focusing on their knowledge, skills, and practices in diabetes management in LMICs, including blood glucose monitoring, lifestyle modification, and prevention of complications. All the

XSL•FO
RenderX

included studies were implemented in a multicomponent nature, wherein they combined training with other components of the intervention, and only one of them (Catley et al [37]) isolated the intervention effect. Thus, the intervention's (training's) effects could not be isolated from the delivery intensities of the interventions reported in the included studies.

- Comparator: standard diabetes care, no intervention, or other training programs unrelated to diabetes served as comparators.
- Outcomes: the primary outcome of interest was the mean change in the glycated hemoglobin (HbA_{1c}) levels of individuals with diabetes managed by the CHWs. HbA_{1c} values were included even when reported as secondary outcomes in a study. The secondary outcomes of interest were other clinical measures, such as random and fasting blood glucose, lipid profile, total cholesterol, and blood pressure (systolic and diastolic). A study was included only when HbA_{1c} levels were reported. HbA_{1c} was chosen as the primary clinical outcome of interest as it is the gold standard recommended by the WHO and the International Diabetes Federation, as it reports the glycemic control of the patient over a 3-month period. However, limiting the analysis to only this outcome would have excluded studies using alternative clinical outcomes, thereby limiting our conclusions regarding the effect of CHW training on diabetes outcomes. Additionally, our research question encompassed diabetes screening along with management, but no eligible studies reporting screening outcomes were identified.
- Study setting: studies conducted in LMICs, as defined by the World Bank, were included (countries with a gross national income per capita between US $1136 and US $4465) [38].
- Exclusion criteria: we excluded studies that did not have CHWs as defined by the WHO; studies that did not train CHWs; studies that did not report HbA_{1c} as an outcome; studies that were not conducted in LMICs; studies that were protocols, reviews, qualitative studies, editorials, and conference abstracts; and studies that were not published in English.

It is important to note that CHW capacity building and service delivery exposure are distinct components of the intervention. Capacity building refers to the structured training interventions provided to the CHWs, with diversity in the curriculum, pedagogy, mode of training, and frequency of training. The service delivery component refers to the activities concentrating on the contact between the CHWs and the patients, such as the type of contact, clinical monitoring, and paraclinical support. Thus, while the capacity building component focuses on CHW competency, the service delivery component focuses on postintervention implementation. Both of these components together represent the effectiveness of the intervention.

Search Strategy

A comprehensive search was conducted across the following databases: PubMed, MEDLINE Ovid, Scopus, EBMR, and CINAHL, spanning the period from January 2000 to April 2024. This period was chosen because the World Medical Association and the WHO published their respective task-shifting guidelines using CHWs around this time [39,40]. The search was limited to studies published in the English language. It used MeSH terms and free text related to CHWs, type 2 diabetes, training, and outcome measures such as HbA_{1c}, random blood glucose, oral glucose tolerance test, and microvascular complications. Additional searches were conducted in gray literature databases, including the WHO International Clinical Trials Registry Platform and ClinicalTrials.gov. The detailed search strategy is provided in Annexure 2 in Multimedia Appendix 1.

Study Selection and Data Extraction

A total of 2 reviewers (AGG and SK) independently screened the titles and abstracts of studies that might meet the inclusion criteria. Data were retrieved for full-text review, and those that fulfilled the inclusion criteria were extracted. A third reviewer (GRB) resolved any disagreements.

Data extraction was conducted using a customized version of the standardized form developed in Covidence (Veritas Health Innovation) software. An initial preliminary round of data extraction was performed with 4 studies each by AGG and SK. Once their interreviewer consistency was confirmed, further extraction was carried out. The data collected included study characteristics (author, year, country, and sample size), details of the intervention (CHW training content, duration, and frequency), and outcome measures (mean change in HbA_{1c}, expressed as a percentage, and secondary outcomes). Additionally, all training-related data points were extracted, including the contents of the training module, mode of delivery, duration, and frequency of training and refresher sessions (if applicable). This was done to assess whether any of these factors were related to our outcome of interest and, if so, to what extent. All data were organized into tables for narrative comparison among the studies.

Quality Assessment

The quality of the included studies was assessed using the second version of the Cochrane Risk of Bias 2 (RoB 2) tool [41] for RCTs and the ROBINS-I tool [42] for nonRCTs. The RoB 2 tool evaluated RCTs across 5 domains: bias from randomization, deviations, missing data, outcome measurement, and reporting. The ROBINS-1 tool assessed nonRCTs across 7 domains: bias from confounding, participant selection, intervention classification, deviations, missing data, outcome measurement, and reporting. Each domain was scored as follows: low risk, some concerns, moderate risk, serious risk, and critical risk. Additionally, the masking of the assessors was not reported in the included studies, so it is not clear whether these studies were (partly) blinded or not.

Data Synthesis and Analysis

The data from all of the included studies were extracted using forms we developed with the Covidence software. All data were exported into Microsoft Excel (Office 2021). The studies were subsequently analyzed narratively, considering the aim and objectives, the inclusion and exclusion criteria, the training characteristics, the outcome values for HbA_{1c}, and the strengths and limitations of each study. We summarized the findings

XSL•FO
RenderX

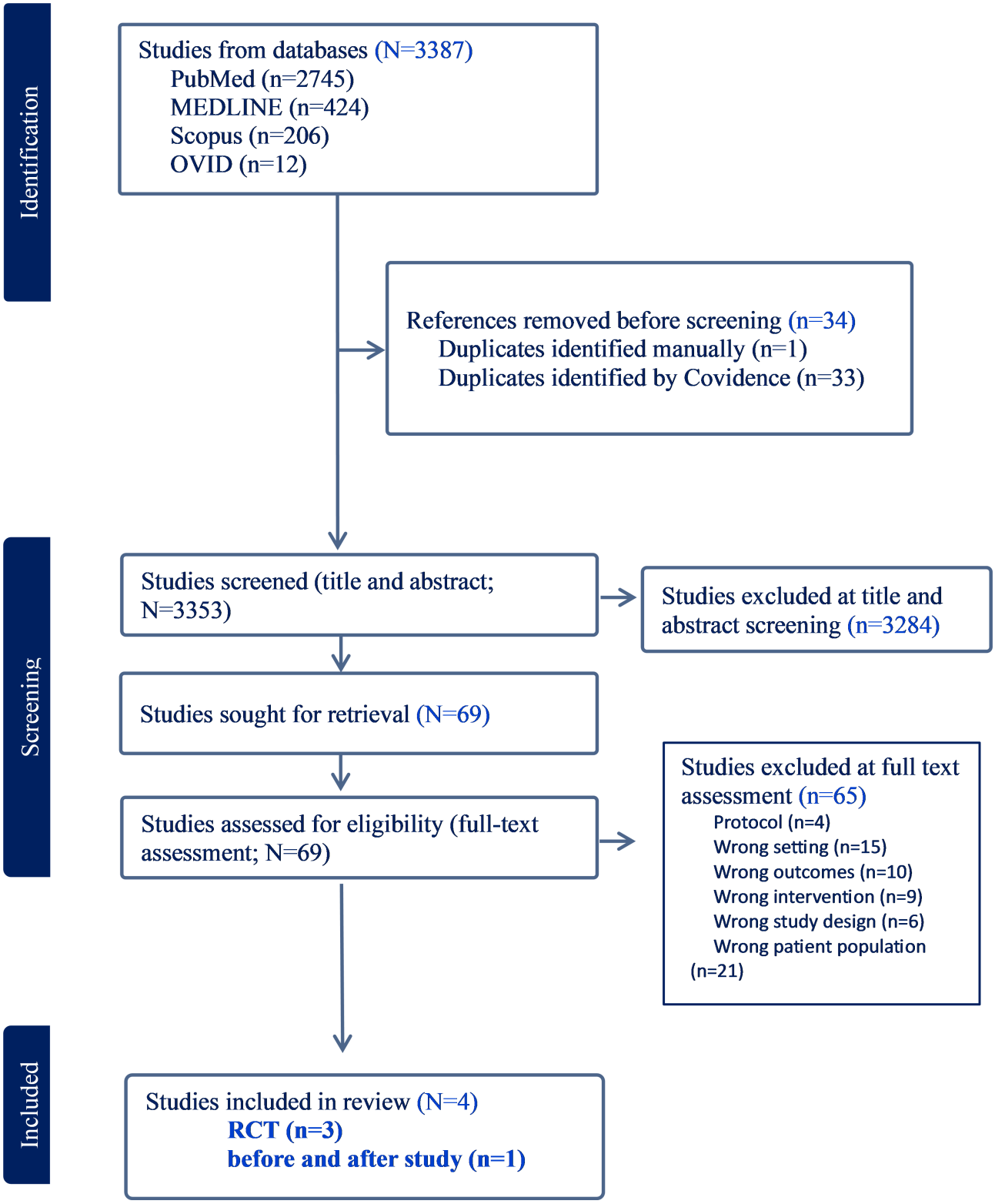
descriptively in text and a comparative table, recording which training features seemed most aligned with improved glycemic control. This approach enabled us to explore thematic links between specific training modalities and HbA_{1c} outcomes without formally combining effect sizes into a single quantitative estimate. A meta-analysis was not performed in accordance with the Cochrane Handbook's recommendation against conducting a meta-analysis with a small number of studies, as the pooled evidence may report unreliable precision and heterogeneity among the studies [43]. The heterogeneity among the studies was conceptual and intervention-based rather than statistical.

Results

We identified 3387 studies for abstract screening, of which 69 were found eligible for full-text review (Figure 1). Upon further screening, 4 studies conducted in LMICs were included, all of which involved the assessment of the impact of training CHWs on the mean reduction of HbA_{1c} levels. However, none of these studies reported any diabetes screening outcomes. Table 1 presents the study characteristics, training characteristics, sample sizes, and primary and secondary outcomes.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart diagram: Interventions involving community health workers training for diabetes in low- and middle-income countries.



XSL•FO
RenderX

Table . Study characteristics of interventions involving community health worker training for diabetes in low-income and middle-income countries.

Study	Study design	Number of CHWs[a] trained	CHW training	Participant characteristics	HbA_{1c}%[b]			HbA_{1c} outcome	Follow-up
					Arm comparison at follow-up	Change within arms	Arm comparison at follow-up		
• Jain et al [44]	• Open-label RCT[c] • 6 months follow-up • Maharashtra, India	• Intervention: 2 new CHWs • Control: NM[d]	• 7-day workshop on communication, adherence, complication recognition, and monitoring of diabetes	• Patients with type 2 diabetes	• Intervention (n=151): 7.63 (SD 2.16) • Control (n=139): 7.64 (SD 1.79) (at 6 mo)	• Intervention: $P<.01$ • Control: $P<.01$ • Statistical test: paired *t* test	• $P=.95$ • Statistical test: ANCOVA adjusting for baseline	• Primary	• 6

XSL•FO
RenderX

Study	Study design	Number of CHWs[a] trained	CHW training	Participant characteristics	HbA_{1c}%[b]				HbA_{1c} outcome	Follow-up
					Arm comparison at follow-up	Change within arms	Arm comparison at follow-up			
Catley et al [37][e]	• Cluster RCT • 7 - 9 months follow-up • Cape Town, South Africa	• NM	• 3-day workshop and 8 weekly half-day experiential sessions on motivational interviewing and behavior-change techniques.	• Patients with BMI≥25 kg/m^2 • support group members receiving health services from local nongovernmental organizations • Intervention (n=240): normoglycemia (n=65, 27.1%), prediabetes (n=106, 44.2%), diabetes (n=69, 28.8%) • Control (n=254): normoglycemia (n=62, 24.4%), prediabetes (n=135, 53.1%), diabetes (n=57, 22.4%).	• Intervention (n=215): 6.32 (SD 1.34); normoglycemia • Control (n=223): mean 6.38 (SD 1.54), normoglycemia (n=64, 28.7%), prediabetes (n=106, 47.5%), and diabetes (n=53, 23.8%).	• Intervention: NM • Control: NM	• Adjusted mean difference: –0.24 (SD 0.08) • *P*=.001 • Statistical test: multilevel ANCOVA fixed effects: arm and baseline value	• Secondary	• 7 - 9
• de Souza et al [45]	• RCT; 3 months follow-up • Conducted in São Pedro, Brazil	• Intervention: 4 new CHWs • Control: 4 new CHWs		• Patients with type 2 diabetes are included in CHW care	• Intervention (n=62): mean 7.9 (SD 1.9) • Control (n=56): mean 8.4 (SD 2.5)	• Intervention: *P*<.001 • Control: *P*<.001			• Primary	• 3

XSL•FO
RenderX

Study	Study design	Number of CHWs[a] trained	CHW training	Participant characteristics	HbA_{1c}%[b]			HbA_{1c} outcome	Follow-up
					Arm comparison at follow-up	Change within arms	Arm comparison at follow-up		
			• Intervention: 4 weekly 60-min diabetes sessions on diabetes definition, lifestyle therapy, pharmacological options, and chronic complications. • Control: 4 classes on tuberculosis, asthma, and contraception. • All CHWs made monthly home visits				• *P*=.13 • Statistical test: Student *t* test (HbA_{1c}), repeated-measures ANOVA, and log-transform for skewed TG[f]		
• Worster et al [46]	• Observational stepped-wedge study with quarterly follow-up for 12 months • Conducted in Chiapas, Mexico	• NM		• Patients (>18 y) diagnosed with and being treated for diabetes and/or hypertension	• N/A[g]	• –0.35% (95% CI –0.90 to 0.20); *P*=.21	• Linear mixed-effects (random intercepts for person and community)	• Primary (for patients with diabetes)	• 12

XSL•FO
RenderX

Study	Study design	Number of CHWs[a] trained	CHW training	Participant characteristics	HbA_{1c}%[b]			HbA_{1c} outcome	Follow-up
					Arm comparison at follow-up	Change within arms	Arm comparison at follow-up		
			• 1 month, 16-session course with monthly refreshers on pathophysiology, motivational interviewing, and home visit logistics. The CHWs received refresher training throughout the 12-month period.						

[a]CHW: community health workers.

[b]HbA_{1c}: glycated hemoglobin.

[c]RCT: randomized controlled trial.

[d]NM: not mentioned.

[e]Catley et al is the only study to report a statistically significant HbA_{1c} difference between the arms [37].

[f]TG: triglycerides.

[g]N/A: not applicable.

Overall, 4 studies involving a total of at least 10 CHWs and around 1000 patients were eligible for inclusion, of which 3 were RCTs [37,44,45] and one was an observational stepped-wedge study [46]. We observed that the training varied in content, delivery, duration, frequency, and settings (Table 1). However, some of the content overlapped with common topics, such as diabetes management, lifestyle modification, and patient education strategies.

Jain et al [44] conducted an open-label RCT in rural Maharashtra, India, with 299 patients with type 2 diabetes: 153 in the intervention arm (trained CHWs) and 146 in the control arm. At the 6-month follow-up, 290 participants responded: 151 from the intervention group and 139 from the control group. A total of 2 CHWs were trained in a 7-day workshop, carried out for home visits and 12 phone reminders over 6 months. Results reported no significant between-arm differences in HbA_{1c} at 6 months (P=.51). Both arms showed HbA_{1c} from baseline (intervention: mean 7.63%, SD 2.16% vs control: mean 7.64%, SD 1.79%; P=.95;Table 1). Similarly, no differences were found for glucose, blood pressure, low-density lipoproteins, high-density lipoproteins, or triglycerides. The RoB 2 assessment indicated some concerns, and the number of CHWs included in the intervention arm was very small (n=2); the number in the control arm was not reported.

The study by Catley et al [37] was a cluster-RCT conducted in Cape Town suburbs, assessing weight loss as the primary

XSL•FO
RenderX

outcome and HbA_{1c} changes as a secondary outcome. Around 28 clusters were randomized between the intervention group (Lifestyle Africa) and the control group. The intervention, Lifestyle Africa, was adapted from the Diabetes Prevention Program, a lifestyle modification program targeting weight loss through motivational interviewing and behavior change methods. Out of 494 enrolled overweight or obese adults, 438 completed the 7 to 9 month follow-up: 215 in the intervention group and 223 in the control group. Among the enrolled participants, 28.8% (69/240) in the intervention arm and 22.4% (57/254) in the control arm had diabetes. The intervention group included 27.1% (65/240) adults with normal glucose metabolism status, 44.2% (106/240) with prediabetes, and 28.8% (69/240) with diabetes; these categories were based on HBA_{1c}. The control group had 24.4% (62/254) adults with normal glucose metabolism, 53.1% (135/254) with prediabetes, and 22.4% (57/254) with diabetes. The CHWs attended a 3-day workshop and 8 sessions featuring 17 videos on motivational interviewing and behavior change. CHWs role-played as mock participants, practicing activities such as weighing participants, using video equipment, and facilitating discussions. Participants in the intervention arm (Lifestyle Africa) experienced a significant reduction in HbA_{1c} compared to the control group (adjusted mean difference –0.24%; *P*=.001). No significant changes were observed in the primary outcomes, including blood pressure, low-density lipoproteins, high-density lipoproteins, triglycerides, and cholesterol levels. Participants in the intervention arm had a higher probability of improving and a lower probability of worsening their glucose metabolism status by at least 1 category (odds ratio [OR]=1.52, 95% CI 1.04-2.22; *P*=.03). The overall RoB 2 assessment indicated some concerns. Therefore, the HbA_{1c} reduction reported reflects a shift in the overall glycemic risk of the overweight adult population in the study and not a trial designed to test the efficacy of diabetes management among patients with diabetes.

The study by de Souza et al [45] was a randomized controlled trial conducted in Porto Alegre, Brazil. The study aimed to measure the percentage change in mean HbA_{1c} between intervention and control groups. A total of 118 patients were randomly allocated to 8 CHWs, with 4 trained in diabetes through 4 weekly 60-minute sessions (covering diabetes definition, lifestyle, medications, and complications) in the intervention arm, which included 62 patients, and 4 in the control arm, which included 56 patients, trained in unrelated health topics such as tuberculosis, asthma, and contraception. All CHWs made monthly home visits. At the 3-month follow-up, there was no statistically significant difference in the percentage of mean HbA_{1c} between the arms (intervention: mean 7.9%, SD 1.9%; control: mean 8.4%, SD 2.5%; *P*=.13). Total cholesterol and triglyceride levels declined more in the intervention group, whereas blood pressure and high-density lipoprotein levels changed similarly in both groups. The overall RoB 2 score indicated high risk, and the number of CHWs in each arm was relatively small (n=4 in each arm).

The study by Worster et al [46] was an observational study conducted in the communities of Chiapas, Mexico. It aimed to study changes in diabetes and hypertension prevalence among patients through CHWs in primary care settings. In total, 149 patients were included, with 73 treated for diabetes. The study followed an observational stepped wedge design over 12 months, with data collection at baseline and every 3 months for 12 months after the implementation of the intervention with newly trained CHWs. In the subgroup of participants with diabetes and a baseline HbA_{1c}≥9%, this value significantly decreased by ~0.96% (*P*=.01), but for all patients with diabetes, the mean decrease (~0.35%) was not statistically significant (*P*=.21). The risk of bias was assessed as serious using the ROBINS-I tool. (The detailed scores with explanations of the various domains for both RoB-2 and ROBINS-I tools are shown in Annexure 3 in Multimedia Appendix 2).

Discussion

Principal Findings

The results of our systematic review show that improvement in HbA_{1c} (at the population level) was moderate and statistically significant in only 1 study by Catley et al [37]. Since HbA_{1c} was not the primary outcome of interest among its sample of individuals with normoglycemia, prediabetes, and diabetes, our review concludes that the evidence for training CHWs in diabetes management in LMICs is limited and, at most, indirect. The limited number of eligible studies, the heterogeneity in training methodologies (ranging from 4-hour sessions to multiday workshops), study designs, and the multicomponent nature of the included interventions limit the interpretation of our results. Therefore, the existing evidence remains inadequate to definitively conclude whether CHW training significantly improves diabetes management across LMICs.

The Catley et al's [37] study evaluated "Lifestyle Africa," an adapted version of the Diabetes Prevention Program, a structured lifestyle modification program that primarily targeted weight loss through motivational interviewing and behavior change techniques. The study involved CHW-patient contact after CHW training, with nearly 29 sessions (17 core and 12 postcore sessions) and a maximum follow-up period of 12 months (varying from 3 to 12 months), thereby providing sufficient time to detect a significant difference in HbA_{1c} levels. However, the study included patients who were normoglycemic, prediabetic, and diabetic, and the change in HbA_{1c} values (a secondary outcome) was recorded for this total population. Thus, the results may not be attributed directly to people with diabetes or generalized to the subgroup of people with only diabetes. The other 2 RCTs (Jain et al [44] and de Souza et al [45]) showed a reduction in HbA_{1c} levels within both arms, but no difference was observed between the study arms. Both studies had shorter CHW-patient interactions (16 contacts and 3 contacts, respectively) and shorter follow-up periods (6 mo and 3 mo, respectively). Meanwhile, the observational study by Worster et al [46] documented a statistically significant change in the percentage of mean HbA_{1c} levels only in patients with HbA_{1c}>9%. However, this could have been due to the regression-to-the-mean effect [47], which may have arisen due to the selection of patients with extreme HbA_{1c} values. In addition, the use of unadjusted statistical tests, like the Student

XSL•FO
RenderX

t test, could have potentially overestimated the actual effect of the intervention. All included studies showed varying bias levels, affecting interpretation. RCTs by Jain et al [44] and Catley et al [37] had some concerns, the RCT by de Souza et al [45] had high bias, and the observational study by Worster et al [46] was rated serious. These ratings highlight the possibility of systematic errors affecting the reliability of and confidence in conclusions derived from the intervention outcomes.

Regarding treatment in the control arms, the participants in the studies by Jain et al [44] and de Souza et al [45] received enhanced usual care, which could have led to changes in behavior and reduced HbA_{1c} levels in the control arms of these studies. Specifically, in the study conducted by Jain et al [44], participants in the control group attended regular physician clinics and interacted frequently with their doctors. In the study by de Souza et al [45], although the CHWs in the control arm were not explicitly trained in diabetes management, they still conducted monthly home visits addressing other health topics (eg, maternal health, immunization, tuberculosis, and asthma), thereby ensuring consistent personal contact. These activities introduce 2 well-documented mechanisms for glycemic improvement. First, repeated contact, measurement, and observation can enhance patients' self-monitoring and adherence to lifestyle advice (the Hawthorne effect), even when the advice is unrelated to diabetes. Second, regular supportive contact (nondiabetes-focused contact) from a trusted CHW or physician can provide nonspecific therapeutic exposure, improving self-efficacy, medication adherence, and problem-solving behaviors, which are known to lower HbA_{1c} [47,48]. Therefore, reductions seen in the control arms are expected to be probable and reasonable, potentially attenuating the intervention effect size.

The outcome variation across studies may result from several factors, primarily due to varying intensity, duration, and content of the training programs for the CHWs. The trial with significant HbA_{1c} benefit (Catley et al [37]) involved a 3-day workshop and 8-day follow-up sessions with ongoing support via videos and SMS text messaging for CHWs spanning 7 - 9 months. The study's training emphasized practical skills, such as motivational interviewing and behavior-change techniques, to enhance patient engagement and coaching. In contrast, the trial by de Souza et al [45] featured a short, single 4-hour training session that emphasized basic diabetes knowledge but lacked ongoing mentorship, leading to no meaningful improvements in glycemic control. These trends may suggest a dose-response relationship, where more intensive, skill-focused training of CHWs improves patient outcomes. These findings imply that CHW training interventions can potentially improve diabetes outcomes, particularly in subpopulations with uncontrolled glycemia; however, consistent and significant effects were not observed across all settings. Published evidence from higher-income countries and from peer-led interventions using CHWs has demonstrated the success of these interventions, advocating for their trial and implementation in LMICs, where a shortage of rural doctors is prevalent. The systematic review by Werfalli et al [49] of peer or CHW-led self-management support in LMICs (where patients managed their diabetes with CHW guidance) reported a statistically significant reduction in HbA_{1c} across studies. Meanwhile, the review by Palmas et al [50] reported modest HbA_{1c} reduction against standard care. The review by Shah et al [51] reported comparable results. Similarly, the DIABLEST trial by Pérez-Escamilla et al [52] and the study by Aponte et al [53] conducted in the United States also reported a statistically significant change in the HbA_{1c} percentage for CHW-led groups compared to the study's control arm. These studies were excluded from our review due to their intervention or location not satisfying our inclusion criteria. Recently, the systematic review by Evans et al [54] reported positive effects across a broader set of studies [54]. Our review differed in scope since we focused on CHW training as the intervention for LMICs and HbA_{1c} as the inclusion criterion. The findings from Evans et al [54] and our review highlighted that, while CHW involvement in diabetes care may be beneficial, the specific and independent contribution of CHW training in LMICs to diabetes care requires further investigation.

The United Kingdom Prospective Diabetes Study [55] showed that a modest reduction in HbA_{1c} levels can be associated with meaningful clinical benefits, such as lowering the risk of future diabetic complications in the eyes, kidneys, and nerves. Even a 1% decrease in HbA_{1c} levels can result in a 37% reduction in combined microvascular complications ($P<.001$), a 21% reduction in risk for any endpoint related to diabetes ($P<.001$), and a 21% reduction in deaths related to diabetes ($P<.001$) [55,56]. Thus, even the approximately 0.3% to 0.5% HbA_{1c} reductions observed in one of our included studies could translate into clinically important benefits. We did not analyze the secondary outcomes, as there was variation in the measurement and reporting of the outcomes across the three RCTs.

Strengths and Limitations

The study has several strengths. This review is one of the first to outline the advantages of training CHWs for improved glycemic control over generic methods used in LMICs. The review also establishes robust methodological rigor, supported by a registered protocol, a PRISMA-guided screening process, and a formal Risk of Bias (RoB) 2 and ROBINS-I appraisal. By integrating a 24-year multidatabase search with pooled and narrative analyses on training intensity and content, this review provides insights relevant to policy and scalable within task-shifting programs.

This study faced limitations, including variability among the included studies in design, intervention, and outcomes. The follow-up periods of all studies varied considerably, ranging from 3 to 12 months. Our focus on LMICs could have narrowed the evidence base while not reflecting the success of CHW training in high-income countries. Thus, the results need to be interpreted cautiously and might not be generalizable. With only 4 trials, testing for publication bias was not performed [57]. Three studies did not isolate the intervention effects of CHW training from other intervention components, so HbA_{1c} changes reflect combined effects, not just training. Only Catley et al [37] isolated the effect of lifestyle coaching statistically. Other studies risked performance and Hawthorne effects due

XSL•FO
RenderX

to additional contact and training, possibly blurring causality. Small sample sizes in Jain et al [44] and de Souza et al [45] may also limit external validity. The scale-up of programs with diverse CHWs may overestimate effects. The implementation of skills gained by CHWs was not reported, which is crucial. Overall, the evidence quality has concerns due to biases and deviations. The certainty of the evidence was low to moderate, with concerns surrounding performance bias that might have introduced Hawthorne effects. The high risk of bias in de Souza et al [45], along with concerns about the randomization process, raises questions about the equal distribution between the arms at baseline. The serious risk of bias in Worster et al [46] due to confounding and deviations from the interventions also raises questions about causal inference. All interventions were multicomponent, so individual effects cannot be separated. Despite extensive searches, unpublished studies might have been missed, risking publication bias.

Implications

While results across RCTs might suggest potentially beneficial effects—despite limitations, with only 1 study showing a significant effect—of introducing CHWs in diabetes care in LMICs, the small number of trials and their limitations preclude the drawing of firm conclusions. Therefore, we suggest that training characteristics, subsequent community interventions, and data collection and analyses must be studied further to develop an optimal training program for CHWs to manage diabetes and monitor the performance of the CHWs at the community level. In addition, such trials should be sufficiently powered, with adequate numbers of both CHWs and patients included. Additional trials, designed to isolate the impact of CHW training from other intervention components, such as service delivery intensity and CHW-patient contact frequency, are necessary to thoroughly establish a link between CHW training and diabetes care in LMICs. Finally, we realize that many studies may not have met our stringent eligibility criteria, and thus, the results cannot be generalized. We therefore propose that future scoping reviews be carried out to address this shortcoming.

This review highlights significant implications for health care practice and policy in LMICs. CHWs are essential for improving health care delivery to underserved groups. Offering them contextually relevant, standardized, and comprehensive training, along with ongoing support and periodic refresher sessions, can help sustain and boost the effectiveness of these interventions over time. Policymakers and health leaders should consider incorporating structured diabetes management training into CHW curricula and conducting regular reviews and refresher courses to maintain their skills. Future research should focus on identifying which CHW training elements have the greatest impact. Larger RCTs with more CHWs and standardized outcomes beyond just HbA_{1c} are necessary to better assess specific component efficacy. Moreover, qualitative studies exploring the experiences and challenges of CHWs in implementing diabetes care can offer valuable insights into improving training programs and support systems, complementing the findings from RCTs.

Conclusions

In conclusion, our review suggests that structured CHW training may be an avenue to improve diabetes outcomes in low-resource settings, but current evidence is insufficient, and more robust scientific evidence is needed before firm conclusions can be drawn. Moreover, its implementation should be accompanied by rigorous evaluation to assess and optimize its impact.

Acknowledgments

The authors did not use any generative artificial intelligence technology for any part of this manuscript.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The data supporting the results of our review were obtained from the study-level data of the included publications, which are presented in the manuscript and the table. The data extraction forms used for the review will be available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: AGG, SK, GRB
Data curation: AGG, SK
Formal analysis: AGG
Investigation: AGG, SK
Methodology: AGG, SK, SV, GRB
Supervision: GVSM, NCS, GRB, OvS
Validation: GVSM, NCS, GRB, OvS
Writing – original draft: AGG
Writing – review & editing: AGG, SK, SV, GVSM, NCS, GRB, OvS

XSL•FO
RenderX

Conflicts of Interest

None declared

Multimedia Appendix 1
Annexure-2: search strategies.
[PDF File, 229 KB - diabetes_v11i1e84508_app1.pdf]

Multimedia Appendix 2
Annexure-3: risk of bias.
[PDF File, 200 KB - diabetes_v11i1e84508_app2.pdf]

Checklist 1
PRISMA checklist
[PDF File, 194 KB - diabetes_v11i1e84508_app3.pdf]

References

1. IDF diabetes atlas. : International Diabetes Federation (IDF); 2025 URL: https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/ [accessed 2026-05-23]
2. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
3. Flood D, Seiglie JA, Dunn M, et al. The state of diabetes treatment coverage in 55 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 680 102 adults. Lancet Healthy Longev 2021 Jun;2(6):e340-e351. [doi: 10.1016/s2666-7568(21)00089-1] [Medline: 35211689]
4. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008 Nov;88(11):1254-1264. [doi: 10.2522/ptj.20080020] [Medline: 18801858]
5. Khunti K, Chudasama YV, Gregg EW, et al. Diabetes and multiple long-term conditions: a review of our current global health challenge. Diabetes Care 2023 Dec 1;46(12):2092-2101. [doi: 10.2337/dci23-0035] [Medline: 38011523]
6. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. Nat Rev Endocrinol 2022 Sep;18(9):525-539. [doi: 10.1038/s41574-022-00690-7] [Medline: 35668219]
7. Rushforth B, McCrorie C, Glidewell L, Midgley E, Foy R. Barriers to effective management of type 2 diabetes in primary care: qualitative systematic review. Br J Gen Pract 2016 Feb;66(643):e114-e127. [doi: 10.3399/bjgp16X683509] [Medline: 26823263]
8. Jacobs B, Ir P, Bigdeli M, Annear PL, Van Damme W. Addressing access barriers to health services: an analytical framework for selecting appropriate interventions in low-income Asian countries. Health Policy Plan 2012 Jul;27(4):288-300. [doi: 10.1093/heapol/czr038] [Medline: 21565939]
9. Long H, Huang W, Zheng P, et al. Barriers and facilitators of engaging community health workers in non-communicable disease (NCD) prevention and control in China: a systematic review (2006 2016). Int J Environ Res Public Health 2018 Oct 26;15(11):2378. [doi: 10.3390/ijerph15112378] [Medline: 30373205]
10. Gudlavalleti MVS, John N, Allagh K, et al. Access to health care and employment status of people with disabilities in South India, the SIDE (South India Disability Evidence) study. BMC Public Health 2014 Nov 1;14:1125. [doi: 10.1186/1471-2458-14-1125] [Medline: 25361570]
11. Kumar V, Singh P. Access to healthcare among the Empowered Action Group (EAG) states of India: current status and impeding factors. Natl Med J India 2016;29(5):267-273. [Medline: 28098080]
12. Beran D. The impact of health systems on diabetes care in low and lower middle income countries. Curr Diab Rep 2015 Apr;15(4):20. [doi: 10.1007/s11892-015-0591-8] [Medline: 25721248]
13. Community health workers: what do we know about them? The state of the evidence on programmes, activities, costs and impact on health outcomes of using community health workers. : World Health Organization, Department of Human Resources for Health; 2007 URL: https://chwcentral.org/wp-content/uploads/2013/07/Community-Health-Workers-What-do-we-know-about-them.pdf [accessed 2026-05-23]
14. Ledikwe JH, Kejelepula M, Maupo K, et al. Evaluation of a well-established task-shifting initiative: the lay counselor cadre in Botswana. PLoS One 2013;8(4):e61601. [doi: 10.1371/journal.pone.0061601] [Medline: 23585912]
15. Oru E, Trickey A, Shirali R, Kanters S, Easterbrook P. Decentralisation, integration, and task-shifting in hepatitis C virus infection testing and treatment: a global systematic review and meta-analysis. Lancet Glob Health 2021 Apr;9(4):e431-e445. [doi: 10.1016/S2214-109X(20)30505-2] [Medline: 33639097]
16. Seidman G, Atun R. Does task shifting yield cost savings and improve efficiency for health systems? A systematic review of evidence from low-income and middle-income countries. Hum Resour Health 2017 Apr 13;15(1):29. [doi: 10.1186/s12960-017-0200-9] [Medline: 28407810]

XSL•FO
RenderX

17. Gyamfi J, Plange-Rhule J, Iwelunmor J, et al. Training nurses in task-shifting strategies for the management and control of hypertension in Ghana: a mixed-methods study. BMC Health Serv Res 2017 Feb 2;17(1):104. [doi: 10.1186/s12913-017-2026-5] [Medline: 28148255]
18. Maria JL, Anand TN, Dona B, Prinu J, Prabhakaran D, Jeemon P. Task-sharing interventions for improving control of diabetes in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Glob Health 2021 Feb;9(2):e170-e180. [doi: 10.1016/S2214-109X(20)30449-6] [Medline: 33242455]
19. Lekoubou A, Awah P, Fezeu L, Sobngwi E, Kengne AP. Hypertension, diabetes mellitus and task shifting in their management in sub-Saharan Africa. Int J Environ Res Public Health 2010 Feb;7(2):353-363. [doi: 10.3390/ijerph7020353] [Medline: 20616978]
20. Joshi R, Alim M, Kengne AP, et al. Task shifting for non-communicable disease management in low and middle income countries--a systematic review. PLoS One 2014;9(8):e103754. [doi: 10.1371/journal.pone.0103754] [Medline: 25121789]
21. Labhardt ND, Balo JR, Ndam M, Grimm JJ, Manga E. Task shifting to non-physician clinicians for integrated management of hypertension and diabetes in rural Cameroon: a programme assessment at two years. BMC Health Serv Res 2010 Dec 14;10:1-10. [doi: 10.1186/1472-6963-10-339] [Medline: 21144064]
22. Balasubramanya B, Isaac R, Philip S, et al. Task shifting to frontline community health workers for improved diabetes care in low-resource settings in India: a phase II non-randomized controlled clinical trial. J Glob Health Rep 2020 Nov 16;4:e2020097. [doi: 10.29392/001c.17609]
23. Anand TN, Joseph LM, Geetha AV, Prabhakaran D, Jeemon P. Task sharing with non-physician health-care workers for management of blood pressure in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Glob Health 2019 Jun;7(6):e761-e771. [doi: 10.1016/S2214-109X(19)30077-4] [Medline: 31097278]
24. Watson SI, Sahota H, Taylor CA, Chen YF, Lilford RJ. Cost-effectiveness of health care service delivery interventions in low and middle income countries: a systematic review. Glob Health Res Policy 2018;3:17. [doi: 10.1186/s41256-018-0073-z] [Medline: 29930989]
25. Vaughan K, Kok MC, Witter S, Dieleman M. Costs and cost-effectiveness of community health workers: evidence from a literature review. Hum Resour Health 2015 Sep 1;13:71. [doi: 10.1186/s12960-015-0070-y] [Medline: 26329455]
26. Lopes SC, Cabral AJ, de Sousa B. Community health workers: to train or to restrain? A longitudinal survey to assess the impact of training community health workers in the Bolama Region, Guinea-Bissau. Hum Resour Health 2014 Feb 11;12:1-9. [doi: 10.1186/1478-4491-12-8] [Medline: 24517103]
27. Tsolekile LP, Schneider H, Puoane T. The roles, training and knowledge of community health workers about diabetes and hypertension in Khayelitsha, Cape Town. Curationis 2018 Mar 26;41(1):e1-e8. [doi: 10.4102/curationis.v41i1.1815] [Medline: 29781697]
28. Wagner J, Keuky L, Fraser-King L, Kuoch T, Scully M. Training Cambodian village health support guides in diabetes prevention: effects on guides' knowledge and teaching activities over 6 months. Int J Behav Med 2016 Apr;23(2):162-167. [doi: 10.1007/s12529-015-9515-x] [Medline: 26438042]
29. Bopape M, Mothiba T, Mutambudzi M, Wens J, Bastiaens H. Baseline assessment of knowledge of home based carers for people with diabetes in a rural village in South Africa: a quantitative study. Open Public Health J 2019 Apr 30;12(1):199-205. [doi: 10.2174/1874944501912010199]
30. Bopape MA, Mothiba TM, Bastiaens H, Wens J. What is the impact of a context-specific training program for home-based carers? an evaluation study. Int J Environ Res Public Health 2020 Dec 11;17(24):9263. [doi: 10.3390/ijerph17249263] [Medline: 33322334]
31. Gudlavalleti AG, Babu GR, Agiwal V, Murthy GVS, Schaper NC, van Schayck OCP. Undesirable levels of practice behaviours and associated knowledge amongst community health workers in rural South India responsible for type 2 diabetes screening and management. Int J Environ Res Public Health 2024 Apr 28;21(5):562. [doi: 10.3390/ijerph21050562] [Medline: 38791775]
32. Stephens JH, Addepalli A, Chaudhuri S, et al. Chronic Disease in the Community (CDCom) Program: Hypertension and non-communicable disease care by village health workers in rural Uganda. PLoS One 2021;16(2):e0247464. [doi: 10.1371/journal.pone.0247464] [Medline: 33630935]
33. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
34. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160. [doi: 10.1136/bmj.n160] [Medline: 33781993]
35. Gudlavalleti AG, Babu GR, Kamalakannan S, Murthy GVS, Schaper NC, van Schayck OCP. Training of community health workers in diabetes lead to improved outcomes for diabetes screening and management in Low- and Middle-Income Countries: protocol for a systematic review. JMIR Res Protoc 2024 Aug 21;13(1):e57313. [doi: 10.2196/57313] [Medline: 39167436]
36. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. : World Health Organization, International Diabetes Federation; 2006 URL: https://iris.who.int/server/api/core/bitstreams/ef6a81ae-5db3-4c5c-9136-c047bd8f8344/content [accessed 2026-05-23]

XSL•FO
RenderX

37. Catley D, Puoane T, Tsolekile L, et al. Evaluation of an adapted version of the Diabetes Prevention Program for Low- and Middle-Income Countries: a cluster randomized trial to evaluate “Lifestyle Africa” in South Africa. PLoS Med 2022 Apr;19(4):e1003964. [doi: 10.1371/journal.pmed.1003964] [Medline: 35427357]
38. World bank country and lending groups. The World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2026-05-23]
39. WMA statement on scope of practice, task sharing and task shifting. WMA (World Medical Association). URL: https://www.wma.net/policies-post/wma-resolution-on-task-shifting-from-the-medical-profession/ [accessed 2026-05-30]
40. Task shifting: global recommendations and guidelines. : World Health Organization; 2008 URL: https://www.unaids.org/sites/default/files/media_asset/ttr_taskshifting_en_0.pdf [accessed 2026-05-23]
41. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
42. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
43. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, editors. Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons, Ltd; 2019:241-284. [doi: 10.1002/9781119536604.ch10]
44. Jain V, Joshi R, Idiculla J, Xavier D. Community health worker interventions in type 2 diabetes mellitus patients: assessing the feasibility and effectiveness in rural central India. J Cardiovasc Dis Res 2018 Oct 12;9(3):127-133. [doi: 10.5530/jcdr.2018.3.29]
45. de Souza CF, Dalzochio MB, Zucatti ATN, et al. Efficacy of an education course delivered to community health workers in diabetes control: a randomized clinical trial. Endocrine 2017 Aug;57(2):280-286. [doi: 10.1007/s12020-017-1352-z] [Medline: 28646377]
46. Worster DT, Franke MF, Bazúa R, et al. Observational stepped-wedge analysis of a community health worker-led intervention for diabetes and hypertension in rural Mexico. BMJ Open 2020 Mar 8;10(3):e034749. [doi: 10.1136/bmjopen-2019-034749] [Medline: 32152172]
47. Ostermann T, Willich SN, Lüdtke R. Regression toward the mean--a detection method for unknown population mean based on Mee and Chua’s algorithm. BMC Med Res Methodol 2008 Aug 7;8:1-9. [doi: 10.1186/1471-2288-8-52] [Medline: 18687143]
48. Shao Y, Liang L, Shi L, Wan C, Yu S. The effect of social support on glycemic control in patients with Type 2 Diabetes Mellitus: the mediating roles of self-efficacy and adherence. J Diabetes Res 2017;2017:2804178. [doi: 10.1155/2017/2804178] [Medline: 28626769]
49. Werfalli M, Raubenheimer PJ, Engel M, et al. The effectiveness of peer and community health worker-led self-management support programs for improving diabetes health-related outcomes in adults in low- and-middle-income countries: a systematic review. Syst Rev 2020 Jun 6;9(1):133. [doi: 10.1186/s13643-020-01377-8] [Medline: 32505214]
50. Palmas W, March D, Darakjy S, et al. Community health worker interventions to improve glycemic control in people with diabetes: a systematic review and meta-analysis. J Gen Intern Med 2015 Jul;30(7):1004-1012. [doi: 10.1007/s11606-015-3247-0] [Medline: 25735938]
51. Shah M, Kaselitz E, Heisler M. The role of community health workers in diabetes: update on current literature. Curr Diab Rep 2013 Apr;13(2):163-171. [doi: 10.1007/s11892-012-0359-3] [Medline: 23345198]
52. Pérez-Escamilla R, Damio G, Chhabra J, et al. Impact of a community health workers-led structured program on blood glucose control among latinos with type 2 diabetes: the DIALBEST trial. Diabetes Care 2015 Feb;38(2):197-205. [doi: 10.2337/dc14-0327] [Medline: 25125508]
53. Aponte J, Jackson TD, Wyka K, Ikechi C. Health effectiveness of community health workers as a diabetes self-management intervention. Diab Vasc Dis Res 2017 Jul;14(4):316-326. [doi: 10.1177/1479164117696229] [Medline: 28330385]
54. Evans J, Ha H, White PT. Evaluating the effectiveness of community health worker interventions on glycaemic control in type 2 diabetes mellitus: a systematic review and meta-analysis. BMJ Open 2025 Jul 15;15(7):e096651. [doi: 10.1136/bmjopen-2024-096651] [Medline: 40664405]
55. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-412. [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
56. Yu PC, Bosnyak Z, Ceriello A. The importance of glycated haemoglobin (HbA(1c)) and postprandial glucose (PPG) control on cardiovascular outcomes in patients with type 2 diabetes. Diabetes Res Clin Pract 2010 Jul;89(1):1-9. [doi: 10.1016/j.diabres.2009.12.009] [Medline: 20494468]
57. Page MJ, Higgins JPT, Sterne JAC. Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, editors. Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons, Ltd; 2019:349-374. [doi: 10.1002/9781119536604]

XSL•FO
RenderX

Abbreviations

CHW: community health worker
HbA_{1c}: glycated hemoglobin
LMIC: low- and middle-income country
PICOS: population, intervention, comparator, outcome, and study setting
PRISMA : Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
WHO: World Health Organization

Edited by I Steenstra; submitted 21.Sep.2025; peer-reviewed by EM Vaughan, G Jeet, L Ruggiero; revised version received 02.May.2026; accepted 04.May.2026; published 10.Jun.2026.

Please cite as:
Gudlavalleti AG, Kamalakannan S, Vanbelle (S) S, Gudlavalleti VSM, Schaper NC, Babu GR, Schayck OCPV
Training Community Health Workers for Diabetes Management in Low- and Middle-Income Countries: Systematic Review
JMIR Diabetes 2026;11:e84508
URL: https://diabetes.jmir.org/2026/1/e84508
doi:10.2196/84508

© Anirudh Gaurang Gudlavalleti, Sureshkumar Kamalakannan, Sophie Vanbelle (S), Venkata Satyanarayana Murthy Gudlavalleti, Nicolaas C Schaper, Giridhara R Babu, Onno C P van Schayck. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 10.6.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Obstacles and Enablers Related to Gestational Diabetes Self-Management: Systematic Review Using the Socio-Eecological Model

Bronya H K Luk, DHSc; Polly H X Ma, PhD; Rosenna Wai Ming Chung, MN; Wendy W Zhang, PhD; William Wing-kuen Lam, DHSc

School of Nursing and Health Sciences, Hong Kong Metropolitan University, Homantin, Hong Kong, China (Hong Kong)

Corresponding Author:
William Wing-kuen Lam, DHSc
School of Nursing and Health Sciences, Hong Kong Metropolitan University, Homantin, Hong Kong, China (Hong Kong)

Abstract

Background: Gestational diabetes mellitus (GDM) requires effective self-management to mitigate associated health risks. A comprehensive understanding of the multilevel factors influencing adherence is essential to designing effective, holistic support strategies.

Objective: This systematic review aimed to synthesize the obstacles and enablers related to GDM self-management across all 5 levels of the socioecological model (SEM).

Methods: A systematic search was conducted across 6 databases (MEDLINE, PsycINFO, CINAHL, PubMed, Cochrane Library, and Web of Science) for literature published from January 2010 to October 2025. Thirty studies (24 qualitative, 4 quantitative, and 2 mixed methods) were included. Data on obstacles and enablers were extracted and synthesized using the SEM as an analytical framework.

Results: The analysis identified 11 key factors across the intrapersonal (eg, knowledge and emotional response), interpersonal (eg, functional support network), organizational (eg, health care access and workplace demands), community (eg, food environment and digital information landscape), and policy levels (eg, funding and economic support). The synthesis reveals how these factors interact across levels, creating systemic challenges such as fragmentation of support and information inequity.

Conclusions: Successful GDM self-management support requires integrated, multilevel strategies that address factors ranging from the individual level to the policy level. This review provides an evidence-based SEM framework to inform the development of comprehensive support interventions, including those leveraging digital health platforms, to improve maternal and neonatal outcomes.

Trial Registration: PROSPERO CRD42023491588; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023491588

(JMIR Diabetes 2026;11:e86767) doi:10.2196/86767

KEYWORDS

obstacles; enablers; self-management; socioecological model; gestational diabetes mellitus; GDM; digital health

Introduction

Background

Gestational diabetes mellitus (GDM) refers to any degree of glucose intolerance with onset or first recognition during pregnancy [1,2]. GDM is a common and consequential health condition during pregnancy, associated with short- and long-term risks for both mother and child. Women diagnosed with GDM face a significant risk of severe health complications such as hyperglycemia, respiratory issues, macrosomia, premature birth, birth injuries, labor dystocia, and even pregnancy-induced hypertension and heart disease [3-6]. Furthermore, their children, in later development, are more likely to experience childhood obesity, early onset of type II diabetes, and an increased risk of cardiovascular disease later in life [7,8]. A cohort study conducted in a diabetes center in Hong Kong revealed that out of 238 women with GDM and postpartum impaired glucose tolerance detected during the postdelivery period, one-fifth (20%) of them developed type II diabetes after an average follow-up period of 4.3 years [9].

GDM currently affects 2% to 10% of pregnant women in the United States, 2% to 6% in Europe, 5% to 8% in Australia, and 10% to 15% in China [10,11]. Asian women, compared to women of many other ethnic backgrounds, are at a higher risk of developing GDM [11,12]. As one of the most common health problems in pregnancy, GDM requires proactive self-management, particularly through lifestyle modifications

XSL•FO
RenderX

in diet and exercise, which can reduce complications for both the mother and infant [1,2,7,13].

Self-management refers to an individual's ability to manage the symptoms, treatments, physical and psychosocial consequences, and lifestyle changes inherent to chronic disease [14,15]. GDM self-management includes self-monitoring of blood glucose, adjusting diet, and increasing physical exercise [7,12].

Self-management skills are also necessary for pregnant women to manage their perception of health conditions that may induce stress, from pregnancy to the postnatal period. Reviews have revealed that self-management resulted in statistically and clinically significant improvements in anthropometric outcomes among women with gestational diabetes [16] and provided them with greater control over their diet and body weight [13]. An interventional study on self-management in chronic disease management demonstrated improved health status, health behavior, and self-efficacy, as well as fewer emergency room visits for participants [17]. Studies report that women diagnosed with GDM face challenges in adapting to their new diagnosis [18-21]. Women often feel frustrated by being diagnosed with the condition and experience a lack of control [21-23].

A study found that categories of factors affecting self-management include personal lifestyle characteristics, health status, resources, environmental characteristics, and the health care system. Factors may interact to affect self-management and may exist on a continuum from facilitator to barrier [24]. Understanding the factors that affect self-management could enhance the evaluation of self-management among pregnant women with gestational diabetes. This understanding could also guide the development of interventions designed to address individual needs and improve health outcomes. These findings suggest that focusing solely on individual responsibility is insufficient; a broader perspective is needed to understand the multilevel system of factors influencing self-management.

Guiding Framework: The Socioecological Model

This review emphasizes the use of the socioecological model (SEM) as a guiding framework for understanding the factors that influence self-management behaviors in women with gestational diabetes. The SEM, widely accepted and used for understanding the health behaviors of individuals, takes into account the dynamic interplay between individuals and their environments as determinants of health-related behavior [25]. This model recognizes that an individual's behavior is influenced by multilevel factors [25,26]. Applied to this review, the SEM provides a framework to systematically identify and analyze the obstacles and enablers of GDM self-management across these levels, from personal experience to policy context.

Aims of This Review

The aim of this systematic review is to comprehensively explore and identify the obstacles and enablers to the self-management of gestational diabetes in pregnant women. It seeks to understand the specific obstacles and enablers at various levels—intrapersonal, interpersonal, organizational, community, and policy—that influence the self-management of this condition. The review will use the SEM to provide a holistic understanding of the factors affecting the self-management of gestational diabetes among pregnant women.

Methods

Data Search Strategy

A comprehensive and reproducible search strategy was developed to systematically identify all relevant literature. The search terms were developed through an iterative process involving a review of key papers on GDM self-management to identify relevant terminology, consultation with database thesauri (eg, MeSH in MEDLINE) to incorporate controlled vocabulary, and the combination of 3 core conceptual categories: GDM (eg, "gestational diabetes mellitus"), self-management (eg, "self-care" and "self-monitoring"), and determinants (eg, "barriers," "facilitators," and "enablers"). Boolean operators (AND, OR) were used to link these concepts. The final search strategy was peer-reviewed by a university research librarian to enhance its sensitivity and specificity. Electronic databases to be systematically searched included MEDLINE, PsycINFO, CINAHL, PubMed, Cochrane Library, and Web of Science. The keywords that were searched included the following: ("gestational diabetes" OR "gestational diabetes mellitus" OR "diabetes in pregnancy" OR "GDM") AND ("self-management" OR "self-care" OR "self-regulation" OR "self-monitoring") AND ("barriers" OR "obstacles" OR "challenges" OR "difficulties" OR "issues" OR "problems" OR "facilitators" OR "motivators" OR "enablers").

The initial search was conducted in February 2024 and was limited to papers published in English between 2010 and the end of 2023. To ensure the review included the most current evidence, the search was updated in October 2025 to cover publications from January 2024 to October 2025. A citation search of the included studies was also performed to identify additional relevant studies.

This study has been registered with PROSPERO (CRD42023491588) and was conducted in adherence to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Checklist 1).

Eligibility Criteria

Studies were included if they met the following criteria: (1) the population comprised pregnant women diagnosed with GDM, excluding studies focused solely on the postpartum period; (2) the study investigated barriers, facilitators, obstacles, enablers, challenges, or motivators related to GDM self-management, which included behaviors such as dietary modification, physical activity, self-monitoring of blood glucose, and medication adherence; (3) the study was conducted in any health care or community setting; and (4) the study design was qualitative, quantitative descriptive, or mixed methods. Review papers, commentaries, conference abstracts, and protocol papers were excluded. Only studies published in English from January 2010 onward were considered, incorporating an initial search up to the end of 2023 and an updated search to October 2025.

XSL•FO
RenderX

Study Selection

A total of 5843 records were identified through the initial database search. After removing 853 duplicates using Covidence (Veritas Health Innovation Ltd), 4990 unique studies underwent title and abstract screening, resulting in 126 reports sought for retrieval. Of these, 119 were successfully retrieved for full-text eligibility assessment. Additionally, 32 records from an updated search (2024 - 2025) and 8 records from citation searching were screened, with 11 and 8 proceeding to full-text assessment, respectively. Following full-text review, 25 studies from the database search, 3 from the updated search, and 2 from the citation search met the inclusion criteria, yielding a total of 30 studies in the final synthesis. The study selection process is detailed in the PRISMA flow diagram in *Results* section. Screening and eligibility assessments were performed independently by 2 reviewers, with discrepancies resolved through discussion and, where necessary, arbitration by a third reviewer.

The initial screening of titles and abstracts was conducted independently by 2 reviewers. The full texts of potentially eligible studies were subsequently assessed for eligibility against the inclusion and exclusion criteria, also independently by both reviewers. Any discrepancies regarding the inclusion or exclusion of studies were resolved through discussion with a third reviewer until a consensus was reached. This same collaborative process was applied during the quality assessment stage.

Quality Assessment

The methodological quality of each included study was appraised independently by 2 reviewers using the Mixed Methods Appraisal Tool (MMAT; version 2018; McGill University) [27]. Based on the study designs identified in our review, only the MMAT checklists for qualitative, quantitative, descriptive, and mixed methods studies were applied; no randomized controlled trials or nonrandomized quantitative studies were included. The complete quality assessment results are provided in Multimedia Appendix 1.

Data Analysis and Synthesis

In this review, the terms “barriers” and “obstacles” are used interchangeably, as are the terms “facilitators” and “enablers,” to reflect the language used across the included literature. Data extraction and synthesis followed a convergent qualitative synthesis design, explicitly guided by the SEM [25] as an analytical framework. The process consisted of 3 stages: extraction, thematic synthesis, and framework analysis.

Stage 1: Data Extraction and Initial Coding

Two reviewers independently extracted all relevant text segments (eg, sentences and paragraphs) pertaining to barriers, facilitators, obstacles, or enablers to GDM self-management from the “Results” or “Findings” sections of each included study. These raw data segments were imported into a standardized spreadsheet alongside key study characteristics (author, year, country, and design). Through an iterative process of discussion and consensus, the reviewers inductively coded these segments, generating a preliminary list of 32 distinct factors (eg, “language barrier,” “physical discomfort,” and “emotional support”).

Stage 2: Inductive Thematic Synthesis

Using a constant comparative method, 2 reviewers iteratively analyzed, compared, and grouped the 32 initial factors based on conceptual similarity and shared meaning. This inductive process collapsed the factors into 11 broader, overarching themes that captured the core phenomena influencing self-management (eg, “knowledge barriers” and “language barriers” were synthesized under the theme “Disease and Self-Management Literacy”). Discrepancies in thematic grouping were resolved through consensus discussion, with a third reviewer consulted for arbitration when needed.

Stage 3: Framework Analysis (Mapping to the SEM)

Each of the 11 synthesized themes was mapped to 1 primary level of the SEM (intrapersonal, interpersonal, organizational, community, and policy). This mapping was performed independently by 2 reviewers using standard definitions of the SEM levels [25]. The guiding principle was to assign a theme to the level at which the factor was most directly experienced or enacted by the woman, or where the most relevant intervention point would logically exist. For instance, “lack of shared family responsibility” was classified at the interpersonal level as a dynamic within the woman’s support network, while “personal time constraints” (a related but distinct factor) was classified at the intrapersonal level as an individual logistical challenge. All mapping decisions were documented, and disagreements were resolved through discussion to ensure consistency.

Integration of Quantitative Findings

Quantitative data from the included descriptive studies (eg, percentages of women reporting a specific barrier) were extracted and synthesized narratively. These findings were not analyzed separately for metasynthesis but were used to triangulate, corroborate, and add magnitude to the qualitative themes during the interpretation phase, providing an indication of the prevalence of certain experiences.

Ethical Considerations

This study does not involve human participants.

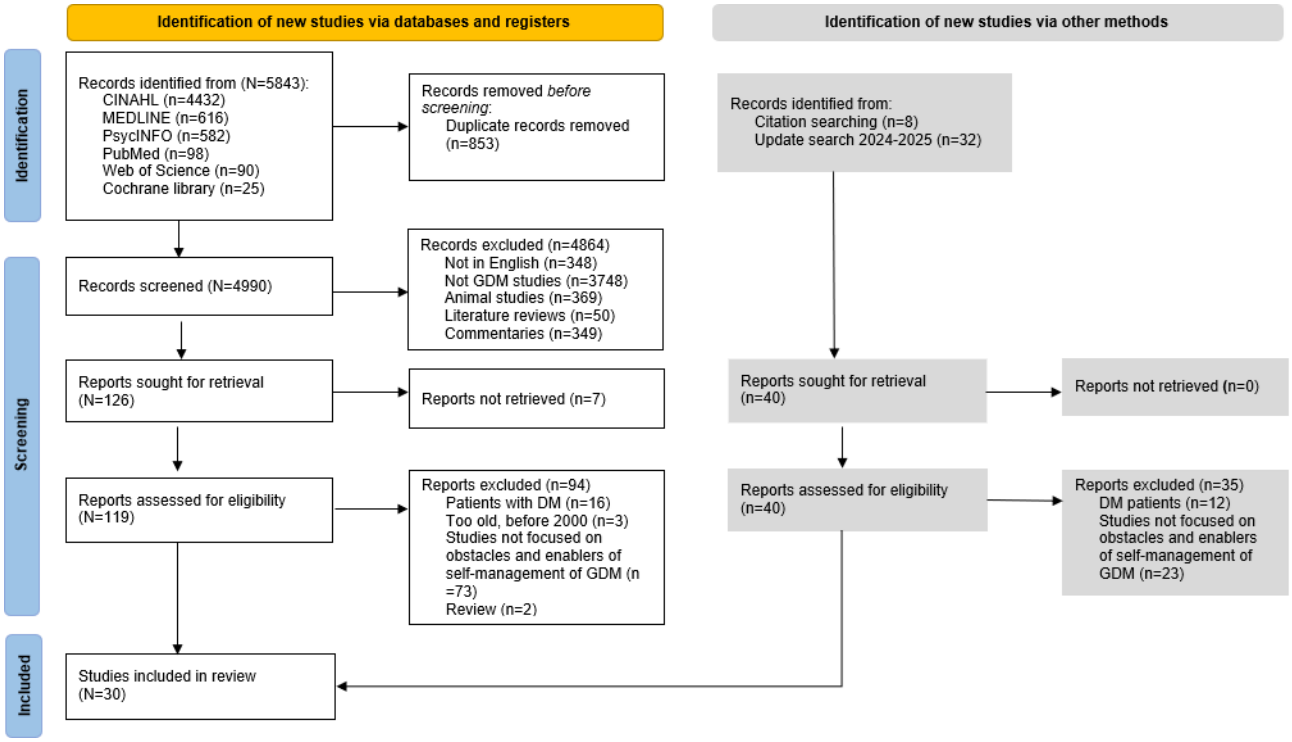
Results

Characteristics of the Selected Studies

This systematic eligibility review process resulted in a final dataset of 30 studies on self-management of GDM, with 24 being qualitative studies, 4 quantitative studies, and 2 mixed methods studies (Figure 1). The identified studies represent a diverse population from 13 countries. Most were conducted in Oceania (n=13), followed by Asia (n=7), North America (n=5), Europe (n=4), and Africa (n=1). The total number of participants in these studies was 2070, with sample sizes ranging from 9 to 564.

XSL•FO
RenderX

Figure 1. The flow diagram depicts different phases of the systematic review. DM: diabetes mellitus; GDM: gestational diabetes mellitus.



The majority of the studies recruited participants through the traditional method of direct contact with pregnant women during clinic visits (n=20). However, newer recruitment methods were also introduced, such as email (n=4), phone (n=3), mobile applications (n=1), and social media (n=2).

In the self-management of GDM, diet modification was the most commonly adopted approach according to 27 studies, followed by exercise (n=23), blood glucose monitoring (n=20), and insulin self-administration (n=3). The characteristics of the selected studies are summarized and listed in Table 1.

XSL•FO
RenderX

Table . Characteristics of the selected studies.

Author, year, and country	Aims	Study design	Participants	Method of recruitment	Type of self-management
Jirojwong et al, 2017, Australia [28]	To explore how Southeast Asian migrant women in Australia experience and manage GDM[a].	Qualitative	19 Southeast Asian migrant women	Clinic patients	Diet, exercise, and blood glucose monitoring
Harrison et al, 2019, Australia [29]	To determine the attitudes of women diagnosed with GDM toward physical activity during pregnancy and to identify the perceived barriers and enablers of physical activity during pregnancy in women with GDM.	Qualitative	27 individuals with diverse ethnicities	Clinic patients	Exercise
de Sequeira et al, 2019, Canada [30]	To explore, among South Asian immigrant women with GDM: (1) their awareness of diabetes education resources, (2) their attitudes toward information from different resources and varying health care providers, and (3) their barriers and facilitators for GDM management recommendations.	Qualitative	13 South Asian immigrant women	Clinic patients	Diet, exercise, and blood glucose monitoring
Skar et al, 2018, Norway [31]	To explore the experiences of women with GDM by controlling their blood glucose values and receiving health and nutrition information using a smartphone app (Pregnant+ app; Philips Digital UK Ltd).	Qualitative	17 individuals with diverse ethnicities	Phone	Diet, exercise, and blood glucose monitoring
Kaptein et al, 2015, Canada [32]	To gain insight into the reactions and experiences of multiethnic women diagnosed with GDM.	Qualitative	19 individuals with diverse ethnicities	Clinic patients	Diet and exercise
Siad et al, 2018, Canada [33]	To document the impact of a diagnosis of GDM and the perceptions of diabetes care among East African immigrant women.	Qualitative	10 East African immigrant women	Clinic patients	Diet, exercise, and blood glucose monitoring
Al Nadhiri et al, 2023, Oman [34]	To examine the predictors, the barriers, and the motivating factors of adherence to the GDM management plan among Arab pregnant women with GDM.	Cross-sectional	164 Arabian	Clinic patients	Diet, exercise, and blood glucose monitoring

XSL•FO
RenderX

Author, year, and country	Aims	Study design	Participants	Method of recruitment	Type of self-management
Khooshehchin et al, 2016, Iran [35]	Aimed at a deeper understanding of women's experiences of gestational diabetes and their perceived needs to inform future lifestyle interventions.	Qualitative	12 Iranian	Clinic patients	Diet and insulin self-administration
Guo et al, 2020, China [36]	To describe the current status of self-monitoring of blood glucose (SMBG) engagement among women with GDM in China and identify its barriers and facilitators.	Mixed methods	323 Chinese women completed the survey and 11 participated in the interview	Clinic patients	Diet, exercise, and blood glucose monitoring
Wah et al, 2019, Australia [37]	To explore the understanding and self-management experiences of gestational diabetes among Chinese migrants.	Qualitative	18 Chinese immigrant women	Clinic patients	Diet, exercise, and blood glucose monitoring
Smyth et al, 2023, Ireland [38]	To explore the attitudes of women with GDM toward these lifestyle changes, specifically diet and exercise. Women were also asked to identify advice that would be useful for other women newly diagnosed with GDM.	Qualitative	32 (28 Irish and European, 4 non-European)	Clinic patients	Diet and exercise
Yee et al, 2015, United States [39]	To evaluate barriers to diabetes self-care during pregnancy in an underserved population.	Qualitative	10 individuals with diverse ethnicities	Clinic patients	Diet, exercise, and blood glucose monitoring
Martis et al, 2017, New Zealand [40]	To explore women's views and experiences in achieving their recommended glycemic treatment targets and to identify potential barriers and enablers.	Cross-sectional	60 European and Asian ethnicity	Mail	Blood glucose monitoring
Kolivand et al, 2018, Iran [41]	To determine the needs of women as an essential first step to formulate a self-care guide fitting the Iranian culture.	Qualitative	13 Iranian	Clinic patients	Diet and exercise
Haghdoost et al, 2019, Iran [42]	To explore the relationship between socioeconomic status and adherence of patients with GDM to medical orders.	Prospective study	230 Iranian women	Clinic patients	Diet, exercise, and blood glucose monitoring

XSL•FO
RenderX

Author, year, and country	Aims	Study design	Participants	Method of recruitment	Type of self-management
Khalil, 2019, France [43]	To understand, from patients' and health care professionals' perspectives, what drives the adoption and diffusion of a telemonitoring solution (myDiabby; MDHC SAS) in a context where telemonitoring activities are still not compensated like traditional follow-ups.	Qualitative	15 women	From the list of patients who had experience of using myDiabby	Diet and blood glucose monitoring
Lawrence et al, 2021, New Zealand [20]	To explore how women diagnosed with GDM perceive dietary recommendations and how this information influences their dietary decisions during pregnancy and beyond.	Qualitative	18 women with a diverse range of ethnic backgrounds	Phone	Diet
Elton, 2022, United Kingdom [44]	To redress a gap in the literature by exploring the intersection between pregnancy and gestational diabetes, and the community that develops therein.	Qualitative	18 women	Online recruitment from a GDMums support group	Diet and exercise
Mukona et al, 2017, Zimbabwe [45]	To explore the barriers to adherence and identify possible solutions to nonadherence to antidiabetic therapy among women with diabetes in pregnancy.	Qualitative	28 women	Clinic patients	Diet and exercise
Kozica-Olenski et al, 2022, Australia [46]	To explore the experiences and acceptability of telehealth use in diabetes in pregnancy care during the COVID-19 pandemic, from the perspectives of pregnant women and their clinicians. The secondary aim was to explore the experiences of pregnant women receiving general maternity care via telehealth during the COVID-19 pandemic.	Qualitative	18 women	Email invitation	Diet, exercise, and blood glucose monitoring
Carolan et al, 2012, Australia [47]	To explore the factors that facilitated or inhibited gestational diabetes self-management among women in a socially deprived area.	Qualitative	15 women with diverse ethnic group	Clinic patients	Diet, exercise, and blood glucose monitoring
Martis et al, 2018, New Zealand [48]	To identify enablers and barriers for women with GDM to achieve optimal glycemic control.	Qualitative	60 women with diverse ethnic group	Email invitation	Blood glucose monitoring

XSL•FO
RenderX

Author, year, and country	Aims	Study design	Participants	Method of recruitment	Type of self-management
Hewage et al, 2020, Singapore [49]	To understand the perception of patients and health care providers of the barriers to GDM management and preferred interventions to manage GDM in an Asian setting.	Mixed methods	216 women (survey) and 15 women (interview)	Clinic patients	Diet, exercise, and blood glucose monitoring
Bandyopadhyay, 2021, Australia [50]	To gain a better understanding of the lived experiences of South Asian women and their experiences of self-management and their health care providers' perspectives of treatment strategies.	Qualitative	23 South Asian women	Clinic patients	Diet, exercise, and blood glucose monitoring
Rasekaba et al, 2021, Australia [51]	To identify the profiles of women accessing care for GDM in a large regional hospital with a rural catchment in Victoria, Australia, as well as gain insight into the views of the women with GDM, clinicians, and IT staff on the acceptability and feasibility of a GDM telehealth in this setting.	Qualitative	9 women	Clinic patients	Diet and blood glucose monitoring
Sabag et al, 2023, Australia [52]	To determine how women with GDM manage their condition and to identify the primary supports and barriers to lifestyle intervention participation.	Cross-sectional	564 women with diverse ethnic group	Online recruited through social media	Diet and exercise
Oxlad et al, 2023, Australia [53]	To explore women's perspectives and experiences concerning how culture impacts the lifestyle management of GDM in women of culturally and linguistically diverse backgrounds.	Qualitative	33 women with diverse ethnic background	Telephoned to invite the penitential participants	Diet
He et al, 2024, China [54]	To gain insight into the experiences of pregnant nurses with GDM in China in terms of their illness, work burdens, and self-care.	Qualitative	9 pregnant nurses	Clinic patients	Diet, exercise, and blood glucose monitoring
Merchant et al, 2025, United States [55]	To examine patient perspectives on social support's role in GDM management during pregnancy and early postpartum.	Qualitative	38 women	Clinic patients	Diet, exercise, blood glucose monitoring, and insulin shot

XSL•FO
RenderX

Author, year, and country	Aims	Study design	Participants	Method of recruitment	Type of self-management
Pham et al, 2025, Australia [56]	To examine women's experiences of GDM online communities on Facebook, their motivations for participation, and perceptions of dynamics within the community.	Qualitative	28 women	Online recruitment through Facebook	Diet, exercise, blood glucose monitoring, and insulin shot

[a]GDM: gestational diabetes mellitus.

Quality Assessment Results

The results, detailed in Multimedia Appendix 1, demonstrate high methodological quality overall. Of the 30 appraised studies, the majority were qualitative (n=24), with quantitative descriptive (n=4) and mixed methods (n=2) studies. The vast majority of criteria were rated "Yes." A small number of "Cannot tell" ratings were present, primarily for criterion 1.5 (qualitative) and criterion 4.4 (quantitative descriptive). No study received more than 2 such ratings, supporting the robustness of the evidence base.

Factors Identified as Obstacles and Enablers

Overview

Across the 30 included studies, a total of 32 distinct barriers and enablers influencing adherence to GDM self-management were extracted. These were synthesized into 11 overarching factors and mapped to 5 levels of the SEM (Table 2). A detailed list of extracted factors is provided in Multimedia Appendix 1. This synthesis offers an evidence-based framework for understanding the multilevel influences on GDM self-management and provides actionable insights to inform the development of targeted interventions aimed at improving adherence and health outcomes for pregnant women with GDM. The following sections present the identified obstacles and enablers at each SEM level, supported by qualitative findings and triangulated with quantitative data from the included studies.

XSL•FO
RenderX

Table . Factors identified as obstacles and enablers to adherence to self-management of gestational diabetes mellitus (GDM) from selected studies.

SEM[a] level and factors	Obstacles	Enablers	Quantitative support
Intrapersonal			
Disease and self-management literacy	Knowledge and language barriers [28,35-39,41,46,47,49,50,53]	Knowledge as empowerment [29,32,33,38,41,44,47,54,56]	• Language barrier: 55% reported not having info in first language [40] • Knowledge is a key factor for adherence [42]
Emotional and cognitive response	Overwhelmed by diagnosis of GDM [20,29,31-33,38,39,50]	Concern for maternal/fetal health [20,28,31,36,38,47], positive attitude [36,54]	• Concern for fetal complications was top motivator for diet (67.1%) and exercise (50%) [34]
Physical capacity and personal logistics	Physical discomfort [28,33,38,45,47,50] and time constraints [28,38,39,47,49,51,54]	—[b]	• Frequent hunger: 61.6% [40] • Lack of time: 71.4% [52]
Interpersonal			
Functional support network	Absence of support from family [20,28,35,39,41,45,47-49,53,54], lack of shared responsibility for family or work commitments [28,32,37,39,45,49,54], and family or friends providing unhealthy food [55]	Emotional and logistical support [28-30,32,35-38,40,41,45,47-51,53-55], and collaborative healthy meal preparation [55]	• Family offering unhealthy food: 38.3% [40] • Husband encouragement: 22.6% - 28.7% [34] • Family commitment: 32.9% - 40.9% [34]
Organizational			
Quality and applicability of health care	Confusing information [28,33,35,36,40,41,47,49,56] and culturally inappropriate advice [20,29,30,32,33,37,50,53]	Comprehensible guidance [28,29,47,49] and supportive patient-provider relationship [20,35-38,41,45,46,49]	• Comprehensible guidance valued by 100% [40] • Support from health care providers valued by 68% [40]
Accessibility and continuity of health care	Fragmented care [37,40,44-46,48,51,56] and hard-to-access services [20,29,30,32,33,37,50,53]	Consistent care [31,51] and technology-supported services [17,31,43,46,51]	• Fragmented care: inconsistent info (16.6%), never seeing same health care provider twice (13.3%) [40] • Online format supported by 62.7% [52]
Workplace demands and accommodations	Shift work or inflexible schedules [54]	Understanding managers and workplace adjustments [54]	—
Community			
Food environment and cultural norms	Limited food choice [28,37,39,44,47,49,53] and cultural food norms [54,55]	—	—
Digital information and peer support landscape	Misinformation of online [30,35]	Trusted online resources [28,30,37,49,54,56] and structured peer support [55,56]	—
Policy			

XSL•FO
RenderX

SEM[a] level and factors	Obstacles	Enablers	Quantitative support
Regulatory and infrastructural systems	Lack of reimbursement or funding for digital health services [43], and lack of workplace accommodation policies [54]	Shared care models [44,49]	—
Economic support systems	High cost of healthy food and devices [45,46]	Financial aid or complimentary services [45,48]	—

[a]SEM: socioecological model.

[b]Not applicable.

Intrapersonal Level

Analysis of the included studies identified 3 key factors at the intrapersonal level: disease and self-management literacy, emotional and cognitive responses, and physical capacity and personal logistics.

Disease and Self-Management Literacy

A significant obstacle was a deficit in knowledge about GDM and its management, compounded by language barriers, which hindered women's ability to seek out and understand necessary resources [28,35-39,41,46,47,49,50,53]. Quantitative studies substantiated this; 33 of 60 (55%) women reported that not having information in their first language was a barrier [40], and another study highlighted the critical role of knowledge in adherence [42]. Conversely, acquiring knowledge was perceived by women as a valuable enabler, supporting their confidence and motivation to engage in self-management behaviors [29,32,33,38,41,44,47,54-56]. Nevertheless, the translation of knowledge into sustained behavioral change remained contingent on additional factors across other SEM levels.

Emotional and Cognitive Responses

The diagnosis of GDM often provoked overwhelming feelings of panic, stress, and being overwhelmed [20,29,31-33,38,39,50]. A primary enabler to counter these emotional challenges was concern for both maternal and fetal health, which motivated adherence to self-management regimens [20,28,31,36,38,47]. Survey data strongly supported this; concern for fetal complications was the most frequently reported motivator for both diet (67.1%) and physical activity (50%) [34]. Furthermore, maintaining a positive attitude was identified as a key factor in sustaining motivation [36,54].

Physical Capacity and Personal Logistics

Physical discomforts, including physical tiredness and frequent hunger, presented substantial obstacles to adhering to diet and exercise recommendations [28,33,38,45,47,50]. This was commonly quantified, with 61.6% (37/60) of women in 1 study reporting frequent hunger as a barrier [40]. Lack of time was the most prevalent barrier in large surveys, reported by 71.4% of women [52].

Interpersonal Level

The functional support network provided by a woman's immediate social circle was a pivotal factor influencing GDM self-management. A significant obstacle was the absence of this support from family and partners, which left women without the necessary encouragement and practical help [20,28,35,39,45,47-49,53]. This challenge was often exacerbated when family or friends provided unhealthy food, creating interpersonal conflict and directly undermining dietary adherence [55]. A related obstacle was the lack of shared responsibility for family and work commitments, which placed the burden of managing GDM solely on the woman amidst her existing duties [28,32,37,39,45,49,54]. Quantitatively, 38.3% (23/60) of women identified being offered unhealthy food by family or colleagues as a barrier [40], while family and child care responsibilities were also frequently cited (32.9% - 40.9%) [34].

Conversely, a robust functional support network served as a powerful enabler. This included emotional and appraisal support—such as encouragement, empathy, and shared accountability—which helped women cope with the stress of their diagnosis and maintain motivation [28-30,32,35-38,40,45,48-50,53-55]. Husband encouragement was a key motivator, reported by 22.6% to 28.7% of women [34]. Tangible, indirect logistical aid was equally critical, with collaborative healthy meal preparation by family members being a key factor in successfully managing dietary changes [28-30,32,35-38,40,45,48-50,53].

Organizational Level

At the organizational level, the structure and delivery of health care services significantly influence self-management. The analysis identified 3 key factors: the quality and applicability of health care, its accessibility and continuity, and workplace demands.

Quality and Applicability of Health Care

A primary obstacle was the provision of confusing or insufficient information about GDM management [28,33,35,36,40,41,47,49,56]. This issue was often compounded by culturally inappropriate advice, particularly regarding dietary recommendations, which reduced the practicality of the guidance for women from diverse backgrounds [20,29,30,32,33,37,50,53]. Enablers included receiving comprehensible guidance [28,29,47,49]—a factor universally valued, with 100% (60/60) of women in 1 study endorsing its importance [40]—and having a supportive patient-provider relationship, which fostered trust and improved adherence [20,35-38,41,45,46,49]. Support from health professionals was also highly valued (41/60, 68%) [40].

XSL•FO
RenderX

Accessibility and Continuity of Health Care

Women reported frustration with fragmented care and a lack of consistent contact with their health care providers [37,40,44-46,48,51,56]. Services were often hard to access due to long waiting times, insufficient consultation durations, and a lack of out-of-office support [20,29,30,32,33,37,50,53]. Quantitative findings specified that inconsistent information (16.6%) and never seeing the same professional twice (13.3%) were reported barriers [40]. A key enabler was the provision of consistent care [31,51]. Furthermore, these accessibility issues were effectively addressed through technology-supported services, such as telehealth, mobile applications, and online platforms, which provided timely feedback, reliable information, and overcame geographical barriers [17,31,43,46,51]. A majority (62.7%) of women in a large survey supported having an online format for lifestyle interventions [52].

Workplace Demands and Accommodations

The demands of a woman's job also presented an organizational-level obstacle, with challenges such as shift work interfering with the ability to maintain consistent meal times and clinical appointments [54]. An enabler identified was supportive managers and workplace adjustments [54].

Community Level

At the community level, broader societal and environmental factors play a significant role. The analysis identified 2 key factors: the food environment and cultural norms, as well as the digital information and peer support landscape.

Food Environment and Cultural Norms

Women faced obstacles related to limited food choices, particularly when navigating meals outside the home and managing the high cost of healthy foods [28,37,39,44,47,49,53]. In addition, they encountered cultural food norms [54,55], where traditional expectations around eating during pregnancy often conflicted with GDM dietary recommendations, creating significant social and personal challenges [28,37,39,44,47,49,53].

Digital Information and Peer Support Landscape

The digital environment presented a dual challenge. A key obstacle was the prevalence of misinformation online, which made it difficult for women to find reliable guidance for GDM self-management [30,35]. Conversely, access to trusted online resources served as a crucial enabler, providing accessible and valuable information [28,30,37,49,54,56]. Beyond information, structured peer support found in organized online groups was also identified as a key enabler, offering a sense of community and shared experience [55,56].

Policy Level

At the policy level, systemic and regulatory factors were found to significantly influence the resources available for GDM self-management. The analysis identified 2 key factors: regulatory and infrastructural systems, and economic support systems.

Regulatory and Infrastructural Systems

A significant obstacle was the lack of reimbursement and funding for digital health services, which limited their integration into standard care and prevented patients from benefiting from them in practice [43]. Another obstacle identified was the lack of workplace accommodation policies for women with GDM [54]. A key enabler to improve care coordination was the implementation of shared care models, which ensure all aspects of GDM management are addressed collaboratively between the woman and her providers, leading to better overall outcomes [44,49].

Economic Support Systems

The high cost of healthy food and monitoring devices posed a substantial financial barrier for many women [45,46]. To mitigate this, the provision of financial aid or complimentary services has been identified as a crucial enabler, ensuring that women have necessary resources to effectively manage their condition [45,48].

Discussion

Principal Findings

This systematic review, using the SEM, reveals that adherence to GDM self-management is not primarily an issue of individual willpower but a complex systemic challenge. Factors enabling or hindering women exist and interact across all 5 levels of the SEM—from personal emotions to national policy. The key contribution of this synthesis is, therefore, not merely the identification of these multilevel determinants but the critical insight into how they interlock to create a context that can either overwhelm or empower women. Moving beyond a simple listing of barriers and facilitators, the following discussion integrates these findings to critically inform the development of effective, holistic support strategies, with particular attention to the role of digital health solutions.

Our findings align with and extend existing literature on chronic disease self-management, which emphasizes the limitations of an individual-focused approach [24]. The significant burdens reported at the intrapersonal level—emotional distress, knowledge deficits, and logistical constraints—are consistent across studies of GDM [29,33]. However, this review underscores that these personal challenges are frequently exacerbated by challenges at other levels. For instance, intrapersonal anxiety and confusion are often directly amplified by organizational factors such as conflicting information from providers or a lack of continuity of care [28,37]. Similarly, the interpersonal obstacle of unsupportive family environments is linked to community-level cultural norms about food and pregnancy [54,55]. This pattern confirms the core SEM tenet that individual behavior is embedded within and shaped by broader systems, a perspective sometimes underemphasized in GDM management research, which focuses predominantly on educational interventions. The interconnected nature of these challenges calls for equally interconnected solutions.

A critical synthesis of the results points to 2 dominant, cross-level themes: fragmentation of support and information inequity. Fragmentation occurs when care is discontinuous

XSL•FO
RenderX

(organizational), social support is absent (interpersonal), and policies fail to coordinate services (policy), leaving women to manage alone. This isolation intensifies intrapersonal stress. Information inequity arises when accessible, trustworthy guidance is scarce. It manifests as a lack of knowledge (intrapersonal), confusing clinical advice (organizational), and an online landscape rife with misinformation (community). These themes are synergistic, creating a vicious cycle that traps women. For example, a woman may receive vague dietary advice from a busy clinic (organizational), seek clarity online only to encounter conflicting information (community), feel increased anxiety and guilt (intrapersonal), and lack a supportive partner to help her navigate these challenges (interpersonal). Effective interventions must, therefore, be designed to bridge these fragments and curate reliable information flows. The relevance of the SEM for understanding GDM-related behaviors is further supported by recent systematic reviews applying the SEM to postpartum physical activity, which similarly identified critical barriers and facilitators at social, organizational, and community levels beyond the individual [57,58].

Consequently, to disrupt these cycles, support strategies must be multipronged and explicitly designed to counteract these systemic failures. Digital health platforms are uniquely positioned to integrate solutions across SEM levels, but their design must be informed by this nuanced understanding. For example, to combat information inequity, a platform must provide centralized, evidence-based, and culturally tailored education (addressing intrapersonal and organizational failures) while incorporating algorithmically vetted content and professional moderation to create a trusted digital community (addressing community-level risks). To address fragmentation of support, features could include shared care plans visible to both patients and providers to improve continuity (organizational), tools to involve and educate family members (interpersonal), and integrated telehealth functions that bypass geographical and time barriers (organizational or community). Crucially, for such tools to be equitable, sustainable, and trustworthy, policy-level advocacy for reimbursement, funding, and robust data security and privacy protection frameworks is essential [43].

These design principles require special consideration for high-risk or underserved populations identified in this review. A critical finding is that challenges related to understanding information should not be reductively classified merely as “language barriers.” For women with limited health literacy—who struggle to comprehend complex medical information even in their primary language—and for working professionals, the health care system’s responsibility extends beyond translation. Providers must prioritize simplifying information, using clear communication, and ensuring cultural relevance to actively improve health literacy. For employed women, such as nurses [54], asynchronous, flexible digital features that accommodate shifting schedules are critical. A one-size-fits-all tool risks perpetuating the very inequities it seeks to solve.

This review has several limitations. Although the methodological quality of the included studies, assessed using the MMAT, was generally high (Multimedia Appendix 1), the preponderance of qualitative studies and those from high-income countries may affect the transferability of findings. The predominance of qualitative studies enriches the understanding of lived experience but limits quantitative generalizability. Furthermore, most included studies were from English-speaking and high-income countries, which may not capture the full spectrum of contextual barriers in low-resource settings and may affect the transferability of findings. Finally, while the SEM provided a valuable organizing framework, the classification of some factors (eg, personal logistics) into a single level is inherently somewhat reductive, as influences often span multiple levels.

Supporting GDM self-management, therefore, demands a systemic view. Achieving meaningful improvement will depend on integrating strategies that simultaneously empower the individual, engage her support network, streamline and humanize health care delivery, foster supportive communities, and advocate for enabling policies. Future intervention research, particularly in digital health, must therefore evaluate not just patient-level outcomes but also how effectively solutions address these interconnected, multilevel determinants to achieve truly comprehensive care.

Implications

The findings of this review underscore the necessity for comprehensive, multilevel strategies to enhance adherence to GDM self-management. These implications target various stakeholders across the socioecological spectrum.

Intrapersonal Level

Health care providers must move beyond simple information delivery to provide education that actively increases GDM knowledge and self-efficacy. This should be coupled with psychological support to help women manage the overwhelming feelings often associated with a diagnosis, as well as practical strategies to mitigate physical challenges like frequent hunger and tiredness, which are significant reported barriers.

Interpersonal Level

As social support is a pivotal enabler, providers should actively facilitate the involvement of partners and family. Educating the support network can transform it from a potential source of conflict (eg, offering unhealthy food) into a source of essential emotional and practical assistance, with spousal encouragement being a key motivator.

Organizational Level

Health care institutions are responsible for ensuring the provision of clear, comprehensive, and culturally sensitive guidelines. To overcome traditional barriers like long waiting times and fragmented care, services should integrate technology, such as telehealth, mobile apps, digital tools, and digital monitoring tools to improve accessibility and continuity—an approach strongly supported by women in need.

Community Level

Community services and public health initiatives should develop and promote accessible resources, including practical food-choice guides and vetted online platforms. Fostering

XSL•FO
RenderX

structured peer support networks can also provide invaluable social support and shared experiential knowledge for women managing GDM.

Policy Level

Policymakers play a crucial role in creating an enabling environment. This includes allocating funding to establish and reimburse digital health services, making them a sustainable part of standard care. Financial subsidies for healthy foods and monitoring devices are essential for equity. Finally, formalizing shared care models through policy can reinforce a collaborative, system-wide approach to GDM management.

Conclusions

In conclusion, this systematic review demonstrates that successful self-management of GDM requires an integrated, multilevel approach informed by the SEM. Moving beyond a focus on individual responsibility, effective support must simultaneously address intrapersonal, interpersonal, organizational, community, and policy factors. By designing interventions and systems that acknowledge and bridge these interconnected levels, health care providers, policymakers, and communities can significantly improve adherence to self-management, ultimately leading to better health outcomes for mothers and their children.

Funding

The authors declared no financial interest was received for this work.

Data Availability

The datasets used and/or analyzed during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

BHKL, PHXM, RWMC, WWZ, and WWKL conceived and designed the study, reviewed and analyzed the selected studies, and read and approved the final manuscript. BHKL drafted the manuscript. PHXM, RWMC, WWZ, and WWKL critically revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mixed Methods Appraisal Tool (MMAT) quality appraisal results.
[DOCX File, 20 KB - diabetes_v11i1e86767_app1.docx]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 38 KB - diabetes_v11i1e86767_app2.docx]

References

1. Charlton M. The evolving management of gestational diabetes. Hong Kong Pract 2016;38(2):53-59 [FREE Full text]
2. HKCOG guidelines: guidelines for the management of gestational diabetes mellitus. : Hong Kong College of Obstetricians and Gynaecologists (HKCOG); 2016 URL: https://www.hkcog.org.hk/hkcog/Download/Guidelines_on_GDM_updated.pdf [accessed 2026-05-05]
3. Feghali MN, Abebe KZ, Comer DM, Caritis S, Catov JM, Scifres CM. Pregnancy outcomes in women with an early diagnosis of gestational diabetes mellitus. Diabetes Res Clin Pract 2018 Apr;138:177-186. [doi: 10.1016/j.diabres.2018.02.004] [Medline: 29427694]
4. Muche AA, Olayemi OO, Gete YK. Effects of gestational diabetes mellitus on risk of adverse maternal outcomes: a prospective cohort study in Northwest Ethiopia. BMC Pregnancy Childbirth 2020 Feb 3;20(1):73. [doi: 10.1186/s12884-020-2759-8] [Medline: 32013909]
5. Natamba BK, Namara AA, Nyirenda MJ. Burden, risk factors and maternal and offspring outcomes of gestational diabetes mellitus (GDM) in sub-Saharan Africa (SSA): a systematic review and meta-analysis. BMC Pregnancy Childbirth 2019 Nov 28;19(1):450. [doi: 10.1186/s12884-019-2593-z] [Medline: 31779584]
6. Sudasinghe BH, Wijeyaratne CN, Ginige PS. Long and short-term outcomes of gestational diabetes mellitus (GDM) among South Asian women—a community-based study. Diabetes Res Clin Pract 2018 Nov;145:93-101. [doi: 10.1016/j.diabres.2018.04.013] [Medline: 29679628]
7. Mensah GP, Ten Ham-Baloyi W, van Rooyen DRM, Jardien-Baboo S. Guidelines for the nursing management of gestational diabetes mellitus: an integrative literature review. Nurs Open 2019;7(1):78-90. [doi: 10.1002/nop2.324] [Medline: 31871693]
8. Veeraswamy S, Vijayam B, Gupta VK, Kapur A. Gestational diabetes: the public health relevance and approach. Diabetes Res Clin Pract 2012 Sep;97(3):350-358. [doi: 10.1016/j.diabres.2012.04.024] [Medline: 22726771]

XSL•FO
RenderX

9. Lee KF, Mak MWH, Lau KO, Chung HHY. Risk of development of diabetes mellitus in Chinese women with persistently impaired glucose tolerance after gestational diabetes. Hong Kong Med J 2011 Jun;17(3):195-201. [Medline: 21636867]
10. Carolan M. Women's experiences of gestational diabetes self-management: a qualitative study. Midwifery 2013 Jun;29(6):637-645. [doi: 10.1016/j.midw.2012.05.013] [Medline: 22877761]
11. Gao C, Sun X, Lu L, Liu F, Yuan J. Prevalence of gestational diabetes mellitus in mainland China: a systematic review and meta-analysis. J Diabetes Investig 2019 Jan;10(1):154-162. [doi: 10.1111/jdi.12854] [Medline: 29683557]
12. Juan J, Yang H. Prevalence, prevention, and lifestyle intervention of gestational diabetes mellitus in China. Int J Environ Res Public Health 2020 Dec 18;17(24):1-14. [doi: 10.3390/ijerph17249517] [Medline: 33353136]
13. Karavasileiadou S, Almegwely W, Alanazi A, Alyami H, Chatzimichailidou S. Self-management and self-efficacy of women with gestational diabetes mellitus: a systematic review. Glob Health Action 2022 Dec 31;15(1):2087298. [doi: 10.1080/16549716.2022.2087298] [Medline: 35867537]
14. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
15. Shepherd E, Gomersall JC, Tieu J, Han S, Crowther CA, Middleton P. Combined diet and exercise interventions for preventing gestational diabetes mellitus. Cochrane Database Syst Rev 2017 Nov 13;11(11):CD010443. [doi: 10.1002/14651858.CD010443.pub3] [Medline: 29129039]
16. Hewage SS, Wu S, Neelakantan N, Yoong J. Systematic review of effectiveness and cost-effectiveness of lifestyle interventions to improve clinical diabetes outcome measures in women with a history of GDM. Clin Nutr ESPEN 2020 Feb;35:20-29. [doi: 10.1016/j.clnesp.2019.10.011] [Medline: 31987117]
17. Lorig KR, Ritter PL, González VM. Hispanic chronic disease self-management: a randomized community-based outcome trial. Nurs Res 2003;52(6):361-369. [doi: 10.1097/00006199-200311000-00003] [Medline: 14639082]
18. Draffin CR, Alderdice FA, McCance DR, et al. Impact of an educational DVD on anxiety and glycaemic control in women diagnosed with gestational diabetes mellitus (GDM): a randomised controlled trial. Diabetes Res Clin Pract 2017 Apr;126:164-171. [doi: 10.1016/j.diabres.2017.02.016] [Medline: 28258027]
19. Hui AL, Sevenhuysen G, Harvey D, Salamon E. Barriers and coping strategies of women with gestational diabetes to follow dietary advice. Women Birth 2014 Dec;27(4):292-297. [doi: 10.1016/j.wombi.2014.07.001] [Medline: 25096174]
20. Lawrence RL, Ward K, Wall CR, Bloomfield FH. New Zealand women's experiences of managing gestational diabetes through diet: a qualitative study. BMC Pregnancy Childbirth 2021 Dec 10;21(1):819. [doi: 10.1186/s12884-021-04297-0] [Medline: 34886814]
21. Neufeld HT. Food perceptions and concerns of aboriginal women coping with gestational diabetes in Winnipeg, Manitoba. J Nutr Educ Behav 2011;43(6):482-491. [doi: 10.1016/j.jneb.2011.05.017] [Medline: 22078771]
22. Catalano PM, McIntyre HD, Cruickshank JK, et al. The hyperglycemia and adverse pregnancy outcome study: associations of GDM and obesity with pregnancy outcomes. Diabetes Care 2012 Apr;35(4):780-786. [doi: 10.2337/dc11-1790] [Medline: 22357187]
23. Draffin CR, Alderdice FA, McCance DR, et al. Exploring the needs, concerns and knowledge of women diagnosed with gestational diabetes: a qualitative study. Midwifery 2016 Sep;40:141-147. [doi: 10.1016/j.midw.2016.06.019] [Medline: 27553869]
24. Schulman-Green D, Jaser SS, Park C, Whittemore R. A metasynthesis of factors affecting self-management of chronic illness. J Adv Nurs 2016 Jul;72(7):1469-1489. [doi: 10.1111/jan.12902] [Medline: 26781649]
25. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
26. Sallis JF, Owen N, Fisher E. Ecological models of health behavior. In: Health Behavior: Theory, Research, and Practice, 5th edition: Jossey-Bass; 2015:43-64.
27. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
28. Jirojwong S, Brownhill S, Dahlen HG, Johnson M, Schmied V. Going up, going down: the experience, control and management of gestational diabetes mellitus among Southeast Asian migrant women living in urban Australia. Health Promot J Austr 2017 Aug;28(2):123-131. [doi: 10.1071/HE15130] [Medline: 27745571]
29. Harrison AL, Taylor NF, Frawley HC, Shields N. Women with gestational diabetes mellitus want clear and practical messages from credible sources about physical activity during pregnancy: a qualitative study. J Physiother 2019 Jan;65(1):37-42. [doi: 10.1016/j.jphys.2018.11.007] [Medline: 30573442]
30. de Sequeira S, Halperin I, Lipscombe LL. Culturally tailored resources for South Asian immigrant women with gestational diabetes: do they work and what's missing? A qualitative study. Can J Diabetes 2019 Dec;43(8):573-579. [doi: 10.1016/j.jcjd.2019.09.007] [Medline: 31787243]
31. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]

XSL•FO
RenderX

32. Kaptein S, Evans M, McTavish S, et al. The subjective impact of a diagnosis of gestational diabetes among ethnically diverse pregnant women: a qualitative study. Can J Diabetes 2015 Apr;39(2):117-122. [doi: 10.1016/j.jcjd.2014.09.005] [Medline: 25512097]
33. Siad FM, Fang XY, Santana MJ, Butalia S, Hebert MA, Rabi DM. Understanding the experiences of East African immigrant women with gestational diabetes mellitus. Can J Diabetes 2018 Dec;42(6):632-638. [doi: 10.1016/j.jcjd.2018.01.013] [Medline: 29914780]
34. Al Nadhiri M, Al Hashmi I, Alaloul F, Al Omari O. Adherence to gestational diabetes mellitus (GDM) management plan among pregnant women in Oman: predictors, barriers, and motivating factors. Diabetes Metab Syndr 2023 May;17(5):102766. [doi: 10.1016/j.dsx.2023.102766] [Medline: 37146357]
35. Khooshehchin TE, Keshavarz Z, Afrakhteh M, Shakibazadeh E, Faghihzadeh S. Perceived needs in women with gestational diabetes: a qualitative study. Electron Physician 2016 Dec;8(12):3412-3420. [doi: 10.19082/3412] [Medline: 28163857]
36. Guo J, Long Q, Li J, et al. Barriers and facilitators of self-monitoring of blood glucose engagement among women with gestational diabetes mellitus in China: a mixed-methods study. Midwifery 2020 Nov;90:102797. [doi: 10.1016/j.midw.2020.102797] [Medline: 32755742]
37. Wah YYE, McGill M, Wong J, Ross GP, Harding AJ, Krass I. Self-management of gestational diabetes among Chinese migrants: a qualitative study. Women Birth 2019 Feb;32(1):e17-e23. [doi: 10.1016/j.wombi.2018.03.001] [Medline: 29693546]
38. Smyth S, Mulligan K, Rutter E, Harrington L, Hatunic M, Higgins MF. Attitudes of women with gestational diabetes toward diet and exercise: a qualitative study. J Matern Fetal Neonatal Med 2023 Dec;36(1):2155045. [doi: 10.1080/14767058.2022.2155045] [Medline: 36599434]
39. Yee LM, McGuire JM, Taylor SM, Niznik CM, Simon MA. “I Was Tired of All the Sticking and Poking”: identifying barriers to diabetes self-care among low-income pregnant women. J Health Care Poor Underserved 2015 Aug;26(3):926-940. [doi: 10.1353/hpu.2015.0073] [Medline: 26320923]
40. Martis R, Brown J, Crowther CA. Views and experiences of New Zealand women with gestational diabetes in achieving glycaemic control targets: the views study. J Diabetes Res 2017;2017:2190812. [doi: 10.1155/2017/2190812] [Medline: 29226155]
41. Kolivand M, Keramat A, Rahimi M, Motaghi Z, Shariati M, Emamian M. Self-care education needs in gestational diabetes tailored to the Iranian culture: a qualitative content analysis. Iran J Nurs Midwifery Res 2018;23(3):222-229. [doi: 10.4103/ijnmr.IJNMR_108_17] [Medline: 29861762]
42. Haghdoost AA, Baneshi MR, Razzaghi A, Noori A. The impact of socio economic factors on the adherence of patients with gestational diabetes mellitus to medical recommendations. Iran J Public Health 2019 Sep;48(9):1690-1696. [Medline: 31700825]
43. Khalil C. Understanding the adoption and diffusion of a telemonitoring solution in gestational diabetes mellitus: qualitative study. JMIR Diabetes 2019 Nov 28;4(4):e13661. [doi: 10.2196/13661] [Medline: 31778118]
44. Elton L. Knowledge, community and care: digital biocitizenship in gestational diabetes. Sociol Health Illn 2022 Nov;44(9):1408-1426. [doi: 10.1111/1467-9566.13516] [Medline: 35972378]
45. Mukona D, Munjanja SP, Zvinavashe M, Stray-Pederson B. Barriers of adherence and possible solutions to nonadherence to antidiabetic therapy in women with diabetes in pregnancy: patients’ perspective. J Diabetes Res 2017;2017:3578075. [doi: 10.1155/2017/3578075] [Medline: 28828389]
46. Kozica-Olenski SL, Soldatos G, Marlow L, Cooray SD, Boyle JA. Exploring the acceptability and experience of receiving diabetes and pregnancy care via telehealth during the COVID-19 pandemic: a qualitative study. BMC Pregnancy Childbirth 2022 Dec 13;22(1):932. [doi: 10.1186/s12884-022-05175-z] [Medline: 36514010]
47. Carolan M, Gill GK, Steele C. Women’s experiences of factors that facilitate or inhibit gestational diabetes self-management. BMC Pregnancy Childbirth 2012 Sep 18;12:99. [doi: 10.1186/1471-2393-12-99] [Medline: 22988897]
48. Martis R, Brown J, McAra-Couper J, Crowther CA. Enablers and barriers for women with gestational diabetes mellitus to achieve optimal glycaemic control - a qualitative study using the theoretical domains framework. BMC Pregnancy Childbirth 2018 Apr 11;18(1):91. [doi: 10.1186/s12884-018-1710-8] [Medline: 29642898]
49. Hewage S, Audimulam J, Sullivan E, Chi C, Yew TW, Yoong J. Barriers to gestational diabetes management and preferred interventions for women with gestational diabetes in Singapore: mixed methods study. JMIR Form Res 2020 Jun 30;4(6):e14486. [doi: 10.2196/14486] [Medline: 32602845]
50. Bandyopadhyay M. Gestational diabetes mellitus: a qualitative study of lived experiences of South Asian immigrant women and perspectives of their health care providers in Melbourne, Australia. BMC Pregnancy Childbirth 2021 Jul 9;21(1):500. [doi: 10.1186/s12884-021-03981-5] [Medline: 34243754]
51. Rasekaba T, Nightingale H, Furler J, Lim WK, Triay J, Blackberry I. Women, clinician and IT staff perspectives on telehealth for enhanced gestational diabetes mellitus management in an Australian rural/regional setting. Rural Remote Health 2021 Jan;21(1):5983. [doi: 10.22605/RRH5983] [Medline: 33478229]
52. Sabag A, Houston L, Neale EP, et al. Supports and barriers to lifestyle interventions in women with gestational diabetes mellitus in Australia: a national online survey. Nutrients 2023 Jan 17;15(3):487. [doi: 10.3390/nu15030487] [Medline: 36771195]

XSL•FO
RenderX

53. Oxlad M, Whitburn S, Grieger JA. The complexities of managing gestational diabetes in women of culturally and linguistically diverse backgrounds: a qualitative study of women's experiences. Nutrients 2023 Feb 20;15(4):1053. [doi: 10.3390/nu15041053] [Medline: 36839411]
54. He J, Wang H, Chen X. Experiences and self-care of pregnant nurses with gestational diabetes mellitus: a qualitative study. BMC Nurs 2024 Jan 11;23(1):33. [doi: 10.1186/s12912-023-01679-x] [Medline: 38212724]
55. Merchant T, DiTosto JD, Gomez-Roas M, et al. The role of social support on self-management of gestational diabetes mellitus: a qualitative analysis. J Midwifery Womens Health 2025;70(5):791-799. [doi: 10.1111/jmwh.13782] [Medline: 40576167]
56. Pham S, Churruca K, Ellis LA, Braithwaite J. Women's views about online communities for gestational diabetes mellitus: a qualitative interview study. Midwifery 2025 Mar;142:104280. [doi: 10.1016/j.midw.2024.104280] [Medline: 39765010]
57. Ioannou E, Humphreys H, Homer C, Purvis A. A systematic review and thematic synthesis of the barriers and facilitators to physical activity for women after gestational diabetes: a socio-ecological approach. Br J Diabetes 2023;23(1):2-13. [doi: 10.15277/bjd.2023.413]
58. Ioannou E, Humphreys H, Homer C, Purvis A. Preventing type 2 diabetes after gestational diabetes: a systematic review mapping physical activity components using the socio-ecological model. Matern Child Health J 2024 Aug;28(8):1354-1379. [doi: 10.1007/s10995-024-03948-w] [Medline: 38829517]

Abbreviations

GDM: gestational diabetes mellitus
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SEM: socioecological model

Edited by A Agarwal; submitted 29.Oct.2025; peer-reviewed by E Ioannou, H Wu; revised version received 12.Mar.2026; accepted 05.Apr.2026; published 01.Jun.2026.

Please cite as:
Luk BHK, Ma PHX, Chung RWM, Zhang WW, Lam WWK
Obstacles and Enablers Related to Gestational Diabetes Self-Management: Systematic Review Using the Socio-Eecological Model
JMIR Diabetes 2026;11:e86767
URL: https://diabetes.jmir.org/2026/1/e86767
doi:10.2196/86767

© Bronya H K Luk, Polly H X Ma, Rosenna Wai Ming Chung, Wendy W Zhang, William Wing-kuen Lam. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 1.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Solutions for Type 2 Diabetes and Prediabetes: Systematic Review of Engagement Barriers, Facilitators, and Outcomes

Ayesha Thanthrige, BSc, MBA; Nilmini Wickramasinghe, PhD

School of Computing, Engineering and Mathematical Sciences, La Trobe University, Plenty Road, Bundoora, Victoria, Australia

Corresponding Author:
Ayesha Thanthrige, BSc, MBA
School of Computing, Engineering and Mathematical Sciences, La Trobe University, Plenty Road, Bundoora, Victoria, Australia

Abstract

Background: Digital health interventions, including artificial intelligence (AI)-driven solutions, offer promise for type 2 diabetes mellitus (T2DM) and prediabetes management through enhanced self-management, adherence, and personalization. However, engagement challenges and barriers, particularly among young adults and diverse populations, persist. Existing reviews emphasize clinical outcomes while neglecting engagement factors crucial to intervention success. This review highlights engagement barriers and facilitators, offering insights into improving digital health solutions for diabetes management.

Objective: The objective of this systematic literature review is to explore the barriers, facilitators, and outcomes of digital health interventions, focusing on the current state of AI applications while including partial AI and non-AI interventions, for managing and preventing T2DM and prediabetes, to inform the development of user-centered, inclusive digital health interventions for diabetes care. Unlike prior reviews, this review aims to inform the development of user-centered, inclusive digital health interventions for diabetes care, with a focus on engagement across various AI interventions and diverse populations.

Methods: A systematic search of PubMed, Scopus, CINAHL, and additional sources was conducted for studies published between January 2016 and October 2025. Eligibility criteria included English-language, peer-reviewed studies focused on digital health interventions for adults with T2DM or prediabetes, reporting engagement, barriers, facilitators, or outcomes. Data were synthesized narratively using thematic analysis, guided by self-determination theory and user-centered design. Quality appraisal was conducted using Critical Appraisal Skills Program, Mixed Methods Appraisal Tool, and AMSTAR-2 tools.

Results: From the 37 studies (14 quantitative, 3 qualitative, 7 mixed-methods, and 13 reviews), interventions comprised 19 AI-driven (eg, chatbots, ML models, and conversational agent or hybrid), 3 partially AI-driven, and 15 non-AI solutions (eg, apps and lifestyle programs), mostly from the USA (n=15). Key barriers to engagement included inadequate personalization (15/37, 41%), environmental constraints (11/37, 11%), cultural and language mismatches (14/37, 38%), and AI-specific concerns (eg, bias and privacy). Facilitators included personalized feedback (19/37, 51%), cultural tailoring (17/37, 46%), user-friendly design, and peer support. AI-driven interventions demonstrated moderate improvements in clinical outcomes (eg, lowering HbA_{1c}, weight loss, and normoglycemia conversion). However, these tools often struggled with keeping users involved and building trust. Non-AI solutions performed similarly but lacked adaptive features.

Conclusions: This review offers novel insights by synthesizing engagement barriers and facilitators across AI and non-AI intervention domains, often neglected in previous studies. It highlights the necessity for testing adaptive, culturally tailored, and user-centered AI interventions to address engagement challenges in T2DM and prediabetes management. Integrating personalization, precision, and value-based care can improve outcomes and scalability. The findings guide the creation of inclusive, AI-driven solutions aligned with self-determination theory and user-centered design principles.

(JMIR Diabetes 2026;11:e80582) doi:10.2196/80582

KEYWORDS

type 2 diabetes mellitus; prediabetes; digital health; artificial intelligence; machine learning; chatbots, engagement; user-centered design

XSL•FO
RenderX

Introduction

Background

Diabetes is a critical global public health concern with significant implications for individuals, health care systems, and economies. The International Diabetes Federation reported 537 million adults with diabetes in 2021, projected to increase to 783 million by 2045 [1]. Therefore, global health care spending on diabetes reached approximately US $966 billion in 2021 [2]. Also, type 2 diabetes mellitus (T2DM) complications, including cardiovascular disease, kidney failure, and neuropathy, exacerbate health care costs, particularly in resource-constrained low-middle income countries (LMICs) [3]. Prediabetes, defined by elevated blood glucose levels below the T2DM diagnostic threshold, affected approximately 541 million adults globally in 2021, with significant increases projected by 2030 [1]. Rising obesity rates, sedentary lifestyles, and poor dietary habits worsen the impact of prediabetes [4]. However, prediabetes represents a critical window for intervention to prevent the progression to diabetes, with an estimated 70% lifetime risk of developing T2DM [5] without lifestyle or pharmacological interventions [6].

Digital health technologies have transformed chronic disease management, such as diabetes, by enhancing self-management, improving adherence, and delivering personalized interventions [7]. Furthermore, artificial intelligence (AI)-driven tools, such as chatbots and machine learning models (ML), provide real-time feedback, predictive analytics, and tailored recommendations for better lifestyle choices [8,9]. Hence, these technologies offer scalable solutions to address diabetes and prediabetes management and prevention across diverse populations [10].

While these interventions have potential benefits for individuals, digital health interventions face significant challenges. Recent studies report that high dropout rates and poor sustained engagement reduce the effectiveness of such interventions [11]. Furthermore, AI-specific challenges, such as data availability, cost considerations, AI algorithm performance, bias, and data privacy, emerge as noteworthy barriers hindering the adoption of AI applications and further complicating engagement in diverse populations [12]. Current systematic reviews of digital health interventions for T2DM and prediabetes primarily emphasize clinical outcomes, such as HbA_{1c} reduction and weight loss, which are critical, but limited attention on engagement barriers and facilitators, which are equally important for achieving these clinical outcomes [9,13,14].

Theoretical Frameworks

Self-determination theory (SDT) provides a robust framework for understanding engagement by emphasizing autonomy (eg, user choice), competence (eg, skill-building), and relatedness (eg, social support) [15]. User-centered design (UCD) principles advocate iterative, user-driven development to ensure usability and alignment with cultural and socioeconomic contexts [16]. Notably, SDT and UCD together guide the development of effective digital health interventions and will allow us to enhance their impact [17,18]. To provide a comprehensive understanding of digital health interventions for T2DM and prediabetes, this review includes both AI-driven and non-AI solutions. Non-AI interventions serve as a baseline to evaluate AI's added value in addressing engagement barriers and enhancing clinical outcomes, enabling a comparison that informs the design of future AI-driven solutions. This review aims to synthesize engagement barriers, facilitators, and outcomes of digital health interventions for T2DM and prediabetes management across diverse populations, using SDT and UCD frameworks. The specific objectives are (1) to identify barriers to engagement in these interventions, (2) to determine facilitators that enhance engagement across diverse populations, and (3) to evaluate the effectiveness of digital health interventions in achieving clinical outcomes

Methods

Study Design and Reporting Guidelines

Systematic literature reviews are usually used to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings [19]. This systematic literature review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [20]. We used the PRISMA checklist, which enhances the quality, reproducibility, and completeness of the review, enabling researchers to assess the validity of the methods and findings. The review protocol was developed and registered with OSF [21] to ensure methodological transparency.

Search Strategy

A comprehensive search strategy was developed for Medline (step 1) and refined through consultation with a university librarian using medical subject headings terms (step 2). Multimedia appendix A (Table S1 in Multimedia Appendix 1) represents the search strategy, combining medical subject headings terms and keywords identified with the Population, Intervention, Comparator, and Outcome framework. Then a range of electronic databases was searched, including PubMed, Scopus, and CINAHL supplemented by hand searching and reference lists. (Table S1 in Multimedia Appendix 2) lists databases searched. The search period was restricted to January 2016 through October 2025 to capture recently developed modern digital health interventions, aligning with the rapid evolution of digital health technologies and the post-2015 surge in AI integration (eg, deep learning breakthroughs) [22,23].

The researchers collaborated to determine final papers for inclusion in review through Covidence, the Cochrane Collaboration's platform for systematic reviews. Inclusion and exclusion criteria for study selection are presented in Textbox 1. In this paper, "Managing" refers to interventions for diagnosed T2DM (eg, self-monitoring, adherence support) [24,25] and "preventing" refers to those for prediabetes to delay onset (eg, lifestyle changes) [26,27]. These categories were applied during full-text review to ensure focus on at-risk or diagnosed adults (

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for selecting studies.

Inclusion criteria

- Studies published in English in peer-reviewed journals
- Studies focused on artificial intelligence-driven or digital health interventions (eg, mobile apps, chatbots, SMS text messaging, and wearables) for type 2 diabetes mellitus (T2DM) or prediabetes management or prevention
- Studies included adults aged 18 years to 75 years with T2DM or prediabetes
- Studies reported effectiveness, engagement patterns, barriers, or facilitators
- Studies used quantitative, qualitative, mixed-methods, or review designs
- Studies published between January 2016 and October 2025

Exclusion criteria

- Non-English studies and nonpeer-reviewed sources (eg, editorials and abstracts)
- Solutions without a digital component (eg, solely pharmacological)
- Studies targeting only children (<18) or older adults (>75) without broader adult data (to focus on broader adult populations, as those >75 often have unique comorbidities and digital literacy issues that require separate review)
- Studies not reporting engagement, barriers, facilitators, or relevant outcomes
- Studies exclusively on type 1 diabetes, gestational diabetes, or populations without T2DM or prediabetes. (Prediabetes populations are included in this review as they represent a critical window for prevention interventions); only populations without T2DM or prediabetes are excluded
- Studies published before January 2016.

Study Selection Process

Two reviewers independently screened titles and abstracts followed by full-text review of eligible studies, with disagreements resolved through discussion. Reference lists of included studies and relevant reviews were manually searched for additional studies. The study identification and selection process was documented in a PRISMA flow diagram for transparency. To avoid overlap from the 13 included reviews, primary studies cited within them were cross-checked against our included primaries and only unique insights from reviews were synthesized narratively.

Quality Assessment

Quality appraisal of included studies was completed by the primary researcher and verified by a second reviewer [28]. Qualitative studies, cohort studies, randomized controlled trials (RCTs), and consensus documents were assessed with the Critical Appraisal Skills Program criteria [29], selecting the appropriate checklist based on study design. Systematic reviews and meta-analyses were appraised using AMSTAR-2 [30]. Mixed-methods and developmental studies were evaluated using the Mixed Methods Appraisal Tool [31]. A scoring system calculated a percentage (number of 'Yes' responses divided by total relevant criteria for the study multiplied by 100) with thresholds defined as high (≥80%), moderate (60% - 79%), and low (<60%) [32]. However, studies were not excluded based on quality appraisal. Quality appraisal was primarily conducted by one reviewer with verification by a second, rather than fully independent dual review, potentially introducing minor subjectivity bias.

Data Extraction

Data were recorded using a standardized form, capturing (1) study characteristics (author, year, country, design, sample size, and demographics), (2) solution details (technology type, duration, theoretical framework, and features), (3) outcomes (categorized as: primary engagement metrics [eg, retention and adherence], secondary behavioral changes [eg, diet and physical activity], clinical [eg, HbA_{1c} and weight]), (4) barriers and facilitators, (5) quality indicators, and (6) qualitative findings on engagement.

Data Synthesis and Analysis

Due to heterogeneity in study designs, populations, and solutions, a narrative synthesis was used [33]. Quantitative data (eg, dropout rates, HbA_{1c} changes) were summarized descriptively and integrated narratively to support qualitative themes, for example, meta-analytic HbA_{1c} reductions from reviews [13] contextualized thematic barriers such as personalization lacks. A thematic analysis was conducted using an inductive approach to identify barriers and facilitators to engagement, guided by SDT and UCD principles. Themes on barriers and facilitators were derived from reported findings in included studies, using inductive coding, with reviewer interpretation guided by SDT/UCD. The process involved (1) data familiarization, (2) initial coding, (3) theme identification, (4) theme refinement, (5) theme definition, and (6) prevalence quantification. Thematic analysis was conducted by the primary researcher, with themes reviewed and refined by a second researcher. Analysis was supported by NVivo (version 20.7.0, QSR International Pty Ltd), which was used to organize and code data. The PRISMA "qualitative synthesis" refers to the narrative thematic approach due to heterogeneity, not a meta-analysis; no quantitative analysis was feasible.

Intervention Classification

In this review, "AI-driven interventions" incorporate digital health tools that leverage AI as a core component (eg, ML models and chatbots) or as an enhancement to existing platforms

XSL•FO
RenderX

(eg, mobile apps with AI features). “Partially AI-driven interventions” incorporate AI components (eg, ML for tailored messaging) alongside non-AI features (eg, manual data entry), distinguishing them from fully AI-driven interventions and non-AI interventions (eg, SMS text messaging and basic mobile apps). Reviews were classified based on the interventions they evaluate.

Ethical Considerations

This study is a systematic literature review and did not involve the collection of primary data, enrollment of human participants, or access to identifiable private information. Therefore, according to institutional and national guidelines, this work did not require Institutional Review Board or Research Ethics Board approval. All data analyzed were derived from previously published, peer-reviewed studies that had obtained their own ethics approval and informed consent as required. No new informed consent was required for this review, as no individual-level or identifiable data were collected, used, or reported. All efforts were made to ensure privacy by using only publicly available summarized findings and by not extracting or presenting any identifiable participant information from the included studies. No participants were recruited for this research and accordingly, no compensation was provided.

Results

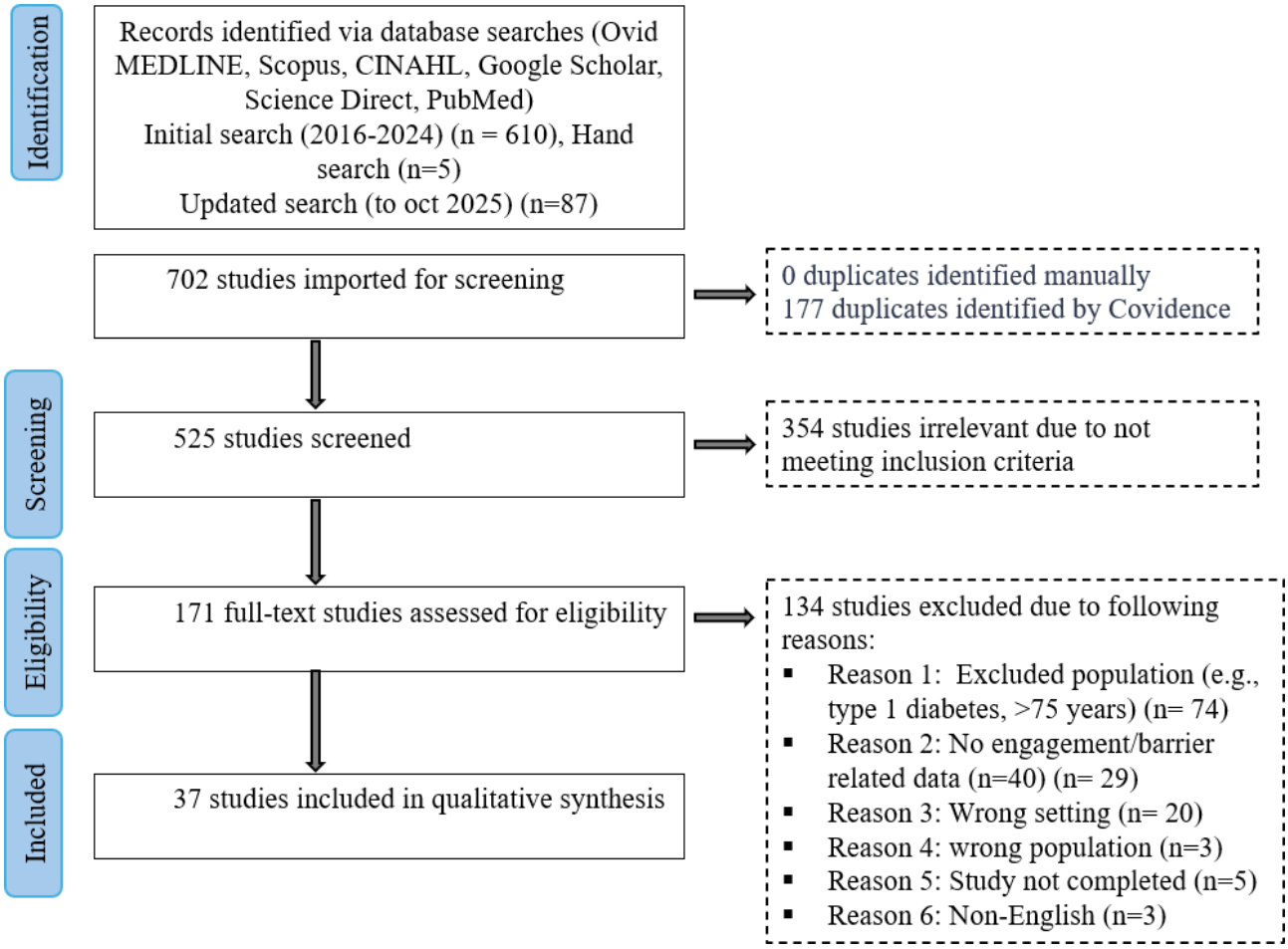
Study Characteristics

Initial search identified (January 2016 to December 2024) 615 studies and were imported into Covidence. An updated search to October 2025 identified 87 additional records. After screening titles and abstracts, 171 full-text articles were evaluated, with 37 studies [8,9,13,14,34-66] meeting the inclusion criteria for the final narrative synthesis (). The majority of the papers were published in 2024 (n=10). The selected studies encompassed diverse populations, including Chinese Americans, Hispanics, Saudi women, and general adult populations, spanning urban and rural settings. Studies originated from various countries, with the majority from the USA/USA-affiliated studies (n=15), and included China, India, Singapore, and Saudi Arabia (Multimedia Appendix 2). Various study designs were observed, including 7 RCTs [8,37,50,57,61,62,64], 7 systematic reviews and meta-analyses [9,13,14,47,54,56,63], 6 narrative, scoping, or other reviews [38,49,58,59,65,66], 5 observational and cohort studies [36,51-53,55], 3 qualitative studies [44,45,60], and 7 mixed-methods studies [39-43,46,48].

Among the 37 studies [8,9,13,14,34-66], 13 are fully AI-based [8,34,35,37,39,40,44-46,50-52,60]: 7 use chatbots, large language models, or conversational agents [8,34,39,40,44-46], 3 use ML models [37,51,60], 2 involve AI-led lifestyle interventions or digital twins [50,52], and 1 uses voice or image recognition [35]. These studies fully leverage technologies, such as chatbots, ML models (many using extreme gradient boosting), conversational agents, and voice or image recognition systems. In total, 8 studies [36,42,43,48,53,57,61,64] are non-AI, consisting of 2 mobile apps, 4 lifestyle programs, 1 SMS text messaging-based intervention, and 1 gamified mHealth application without AI components, relying instead on traditional methods such as lifestyle interventions and conventional digital health tools. Three studies [41,55,62] are partially AI-based, combining AI features with human-led or manual components, a mobile app with tailored messaging and a provider portal, an AI-powered app with automated cues plus dietitian chat, and automated insulin titration systems. The remaining 13 studies[9,13,14,38,47,49,54,56,58,59,63,65,66] are reviews, of which 6 focus on AI applications and 7 address non-AI or broader digital health interventions. Figure 1 details the PRISMA flowchart summary.

XSL•FO
RenderX

Figure 1. Study identification and selection process. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Quality Assessment

Table S1 reports the quality appraisal scores for the 37 [8,9,13,14,34-66] studies, assessed using standardized tools, such as Critical Appraisal Skills Program, Mixed Methods Appraisal Tool, and AMSTAR-2, reflecting rigorous evaluation of methodological quality, with scores ranging from 75% to 100%. No studies were excluded based on quality, underlining the overall rigor of the included research.

Thematic Analysis

Thematic analysis was guided by self-determination theory and UCD principles. Table 1 summarizes themes, subthemes, and findings of digital health interventions.

XSL•FO
RenderX

Table . Themes, subthemes, and findings for digital health interventions in type 2 diabetes mellitus and prediabetes.

Theme	Summary of findings in studies
Barriers to engagement	Factors affecting user engagement and sustainable use of digital health interventions for T2DM[a] and prediabetes, impacting SDT[b] elements autonomy, competence, and relatedness. High dropout rates (6.4% - 35.5% across 14/37 studies) served as a quantifiable indicator of low engagement, primarily due to low-risk perception, inadequate personalization, and lack of motivation [34].
Inadequate Personalization	Found in 15/37 studies (41%), reducing motivation due to generic content or lack of cultural sensitivity includes generic meal plans in ChatGPT responses [34], limited Chinese-specific dietary advice [35], nonindividualized calorie, targets in a formula diet RCT[c] [64], and poor demographic adaptation in ML models [51,60].
Low risk perception	Reported in 4/37 studies (11%), low perceived susceptibility reduced relevance and early adherence particularly among younger adults as users saw interventions as irrelevant [36,37].
Environmental constraints	Present in 11/37 studies (30%), regional health care system variations [38], inadequate access to advanced devices in low-income countries [39], limited smartphone/internet access [9,40], affordability challenges and poor digital literacy [41,42], and infrastructural issues in low-resource settings [57] led to reduced adoption and increased dropout, especially in rural populations [26].
Cultural and language barriers	Found in 14/37 studies (38%), include English-only interfaces [43,44], limited non-English conversational agents [45], limited localization [40] cultural mismatches in content [8,41,42] and culturally insensitive dietary recommendations [34,35] reduced engagement.
AI[d]-specific barriers	Reported in 5/37 studies (14%), include potential for fabricated information [44], limited intent recognition, nonadaptive rules-based systems [8], and inconsistent responses [46], compromising accuracy and personalization in AI-driven interventions.
Socioeconomic barriers	identified in 13/37 studies (35%) include limited smartphone access [47], low digital literacy [42], and restricted access to care [58], hindering adoption and effectiveness in resource-constrained populations.
Facilitators of engagement	Factors enhancing user engagement and adherence, supporting autonomy, competence, and relatedness (SDT) and aligning with UCD[e] principles.
Cultural and linguistic tailoring	Reported in 17/37 studies (46%), enhanced relatedness, by culturally adapted carbohydrate tracking [43], multilingual support with local accents [45], Persian food databases [41], Hispanic-focused soccer programs [36], Arabic WhatsApp peer groups [61], Chinese American web-based DPPs[f] [48] and Chinese-specific dietary recommendations [35].
Personalized and adaptive feedback	Found in 19/37 studies (51%), improved motivation through AI-driven adaptivity or gamification, includes use of personalized AI interactions and conversational empathy [9,14], AI-led DPP [50], personalized, adaptive or AI-powered feedback [35,51,52] and non-AI personalized communication or program choice [53].
User-friendly design	Identified in 5/37 studies (14%), include simple interfaces with no login [51], voice-based interactions [8], automated reminders [9], user-friendly WeChat mini-program [35] and multi-platform access [40], enhancing usability and engagement.
Peer support and social features	Observed in 9/37 studies (24%), include community-driven co-production [60], social support via WhatsApp [36], private Facebook groups [48], family involvement in culturally tailored DPPs [36], buddy systems in hybrid apps [37], and community features in apps [54], enhancing engagement.
Telemonitoring and real-time feedback	Found in 10/37 studies (27%), include continuous glucose monitoring and Fitbit integration [51], real-time data portals [8], real-time monitoring [67], automated alerts through AI systems [50] and cloud-based provider portals [41], built competence and accountability via real-time tracking.

XSL•FO
RenderX

Theme	Summary of findings in studies
Clinical and behavioral outcomes	Measurable impacts of digital health interventions on health outcomes and user behaviors, critical for evaluating effectiveness.
HbA^{g}_{1c} reduction	HbA_{1c} reduction outcomes in diabetes management include a 0.3% decrease with chatbots [9], 0.39% with mobile apps [47], 1.8% with digital twin tech [52] and 1.6% with telehealth interventions [56], improving glycemic control.
Weight loss	Weight loss outcomes in diabetes management include 1.3 kg with chatbots [14], 10.6% with app engagement [55], 7.3% at 6 months [53], up to 6.5 kg in multi-strategy DHIs[h] [56] and 5.9 kg with lifestyle interventions [64].
Normoglycemia conversion	Normoglycemia conversion outcomes in diabetes management show a 50% conversion rate with a low-carbohydrate diet intervention versus 31% with lifestyle alone [64], indicating enhanced reversal of prediabetes through technology-supported dietary strategies.
Improved physical activity	Improved physical activity outcomes in diabetes management include increased step-goal achievement with chatbots [14], enhanced VO2[i] max and agility [36], and improved activity with app-based tracking [47].
Enhanced dietary management	Enhanced dietary management outcomes in diabetes management include improved diet with chatbots [14], 96.43% acceptable ketogenic diet responses [35], and better energy intake with lifestyle interventions [68].
Increased user engagement and adherence	Increased engagement/adherence as primary outcomes, include AI nutrition system with 96.43% valid ketogenic diet recommendations [35], culturally adapted meal adherence [43,48], improved food intake and energy management in mobile and lifestyle programs [47], 89.3% data logging with voice-based AI [8], 65% program completion [53], high chatbot acceptance [9], and 85% retention [48].

[a]T2DM: type 2 diabetes mellitus.
[b]SDT: self-determination theory.
[c]RCT: randomized controlled trial.
[d]AI: artificial intelligence.
[e]UCD: User-centered design.
[f]DPP: diabetes prevention program.
[g]HbA_{1c}: hemoglobin A1c.
[h]DHI: digital health intervention.
[i]VO2: volume of oxygen.

Barriers to Engagement

The thematic analysis identified key barriers such as low-risk perception, inadequate personalization, environmental constraints, cultural or language mismatches, AI-specific concerns (eg, bias, privacy), and socioeconomic barriers. Dropout rates (6.4% - 35.5% across 14 studies [8,34,36,48-53,55,57,61,62,64]) served as an indicator of low engagement, primarily due to underlying causes such as inadequate personalization and motivational lacks, rather than a standalone theme. For example, declining app usage linked to repetitive content [51] and declining motivation over time [34,36], suggesting gamification as a solution, reinforced by generic features in non-AI reviews [41,56].

Declining use over time was frequently attributed to repetitive content and inadequate personalization, both of which reduced users' sense of competence and motivation. For instance [55], reported declining nBuddy app usage, while [35] noted inconsistent AI-driven dietary advice due to limited personalization. Additional studies, such as [49] and [36], highlighted similar challenges with maintaining user motivation over time [34], noted poor sustained engagement from lack of follow-up, and [36] reported declining physical activity post-intervention. Studies suggested gamification, social incentives, and adaptive feedback as remedies to sustain interest.

Inadequate personalization, identified as a critical barrier in 15 [34,35,37,39-41,46,47,50-52,55,57,60,64] of 37 studies[8,9,13,14,34-66] (41%), weakened users' sense of competence and relevance. Static feedback, generic goal-setting, and nonadaptive content led to disengagement. Examples include generic dietary advice in nontailored chatbots [40], poor demographic adaptation in ML models [39,51], and limited personalization in traditional programs [47,69]. Although AI-driven interventions with adaptive algorithms [35,52] demonstrated improved engagement and outcomes, they still exhibited constraints in fine-grained individualization. Several studies emphasized the importance of tailoring content to user preferences, cultural context, and progress level. Examples include generic meal plans in ChatGPT (OpenAI) responses

XSL•FO
RenderX

[34], limited Chinese-specific dietary advice [35], and nonindividualized calorie, targets in a formula diet RCT [64].

Low-risk perception, identified in 4 studies, [34,36,37,49] especially among younger adults, undermined intrinsic motivation and autonomy, leading to disengagement. Participants often viewed interventions as irrelevant to their immediate health needs. Several lifestyle-based, non-AI studies [34,36,49] reported decreased participation due to perceived low personal diabetes risk and lack of urgency.

Environmental constraints, noted in 11 [9,26,38-42,48,53,57,61] out of 37 studies [8,9,13,14,34-66] (30%), included socioeconomic barriers, such as limited smartphone access, and technical challenges, such as complex interfaces, affordability, and accessibility. For instance [42], highlighted technology unfamiliarity among older adults, and [48] reported navigation difficulties. Other studies, including [53] and [61], emphasized socioeconomic barriers, such as transportation or resource limitations, which further hindered engagement.

Cultural/language barriers, in 14 [8,34-36,40-45,47,48,53,61] out of 37 studies[8,9,13,14,34-66] (38%), reduced engagement in diverse populations due to noninclusive content. For example [43], reported an English-only DiaFriend app, and [40] noted an Italian-only AIDA chatbot, and [68] reported high dropout rates (not quantified) in an Arabic-only WhatsApp program, while culturally tailored interventions showed better retention, such as [48] with 15% dropout in a Chinese American web-based diabetes prevention program (DPP). These issues hinder accessibility and personalization, particularly for non-English-speaking and culturally diverse populations, reducing intervention effectiveness.

AI-specific barriers, in 5 [8,40,44-46] out of 37 studies [8,9,13,14,34-66] (14%) included potential for fabricated information [44], constrained input and miscommunication risks [45], limited intent recognition with 9% misclassification [40], and nonadaptive rules-based systems [45], compromising trust, accuracy, and personalization in AI-driven interventions. Socioeconomic barriers, in 13 [36,37,41-43,47,48,53,55,57,58,61,64] out of 37 studies [8,9,13,14,34-66] (35%), involved limited smartphone access [57], resource constraints and access issues [58], limited digital literacy and technology unfamiliarity among older adults [42], smartphone literacy requirements [41], and cost of devices and data.

Facilitators of Engagement

Engagement facilitators were identified in studies reporting high retention or engagement. Cultural tailoring, in 17 [9,14,35,36,40-43,45,48,51,54,55,58,60,61,66] out of 37 studies [8,9,13,14,34-66] (46%), enhanced relatedness through relevant content, as seen in [43] with the culturally tailored app and [36] with a Hispanic-focused program, Portuguese American carbohydrate tracking [43] and [48] with Chinese American web-based DPP (85% retention).

Personalized feedback, in 19 studies [8,9,14,34,35,37,40-42,44-46,50-53,55,57,62] (51%), sustained motivation via adaptive features or gamification. Examples include [8] with voice-based conversational AI achieving 82.9% adherence and 89.3% data logging [50], with AI-powered adaptive interventions showing 93.4% initiation and 63.9% completion rates, and [52] with digital twin technology showing 50.7% diabetes remission rates. User-friendly design, in 5 studies [8,9,35,40,43] (14%), improved accessibility with intuitive interfaces [43]. featured simple interfaces with no login requirements [8], used voice-based interactions for ease of use [40], provided multi-platform access (Telegram [Telegram Messenger LLP], website, Alexa [Amazon]), and [35] offered a user-friendly WeChat (Tencent) mini-program. Peer support, in 9 studies [36,37,48,54,55,58,60,61,66](24%), fostered community engagement, as evidenced by [48] with a Facebook (Meta) group and [61] with WhatsApp support groups [36], with social support via WhatsApp and family involvement, and [37] with a buddy system, enhancing sustained engagement. Telemonitoring and real-time feedback (10 out of 37 [27%]) enhanced engagement via continuous glucose monitors (CGM) or Fitbit integration [51], telemonitoring with scales and pedometers [45], and cloud portals [41], enabling timely intervention adjustments and competence-building.

Clinical and Behavioral Outcomes

Thematic analysis, guided by SDT and UCD, reveals that AI-based interventions for T2D self-management foster autonomy and competence, yielding significant clinical behavioral outcomes, including HbA_{1c} reductions (0.19% - 1.8%), improved diet and physical activity adherence, and weight loss (0.8% - 10.6%) [9,47,55]. Culturally tailored tools and voice-based AI enhance relatedness and engagement, supporting glycemic control [8,43]. Non-AI interventions, such as lifestyle programs, contribute similarly but lack adaptive personalization, emphasizing AI's potential to address SDT-driven motivation gaps [36].

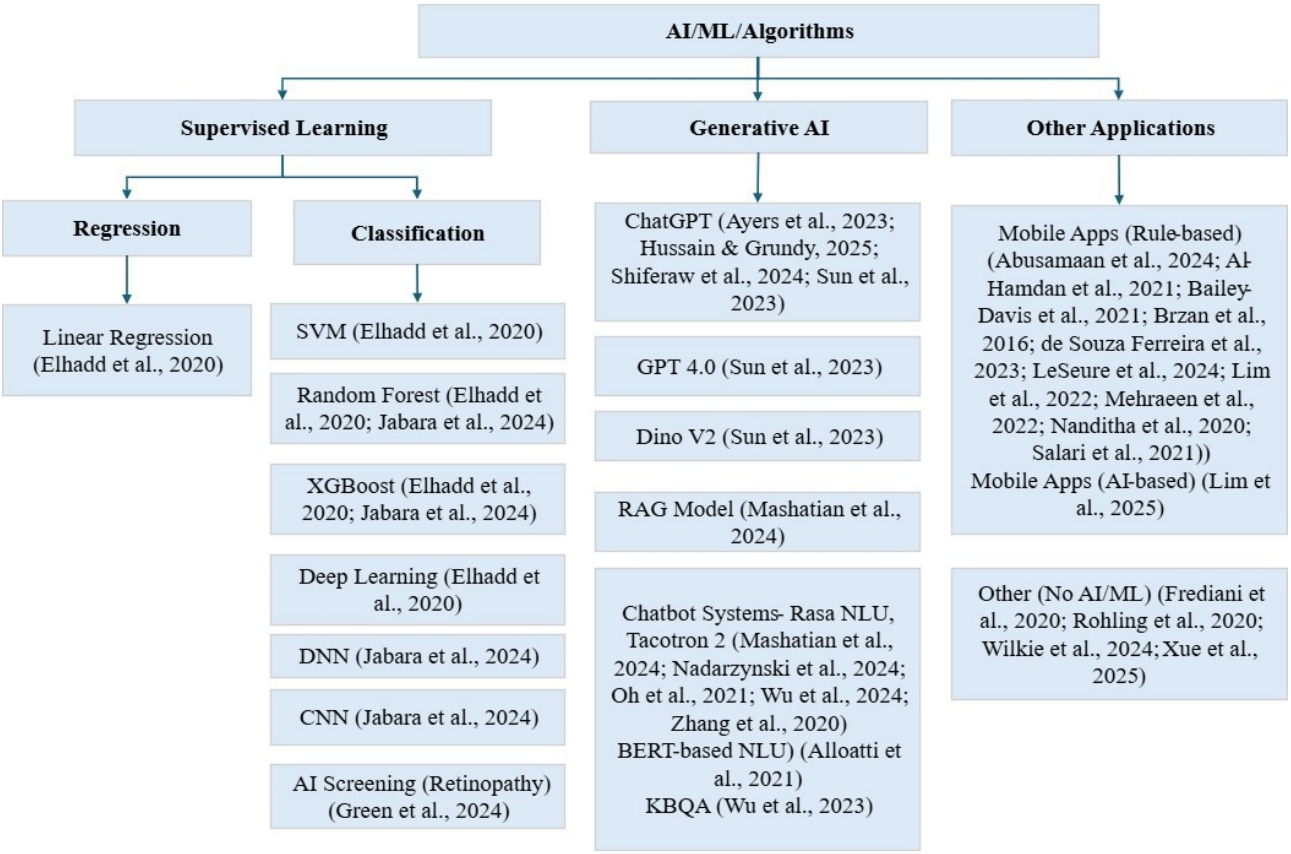
The Application of AI in Diabetes Care

The application of AI in diabetes care, as evidenced by 19 AI-based studies [8,9,14,34,35,37-40,44-46,49-52,59,60,65] among the 37 analyzed, has shown significant potential to enhance engagement and clinical outcomes through advanced methodologies (Figure 2). Quantitative results from a small subset of trials demonstrated moderate-to-high engagement (63.9% - 93.4%) and relatively low dropout (6.6% - 17.9%) compared with non-AI interventions [8,50,52]. For instance, an AI-powered digital twin intervention achieved a 1.8% reduction in HbA_{1c} and 4.8 kg weight loss over one year with 6.6% attrition [52], while a voice-based AI assistant for insulin titration reported 82.9% adherence and faster dose optimization relative to standard care [8]. ML models such as extreme gradient boosting achieved high accuracy in predicting glucose variability (R^2=0.837) [51] and enabled precise screening and complication detection using digital biomarkers from sensors, such as electrocardiograms and photoplethysmography (eg, IDx-DR, an AI-based diabetic retinopathy screening tool: 87.2% sensitivity, 90.7% specificity) [59]. Large language models (LLMs), including ChatGPT and GPT-4, delivered tailored dietary advice, with [35] reporting 74.5% accuracy on the Chinese Registered Dietitian Exam and 96.43% of ketogenic diet responses rated acceptable or excellent, though limited by inconsistent recommendations for Chinese-specific foods.

XSL•FO
RenderX

Conversational agents, such as AIDA [40], reached approximately 4000 unique users with 91% intent recognition accuracy, while AMANDA [45] offered multilingual support with a Singaporean-accented text-to-speech feature, achieving high usability (System Usability Scale score=80.625) and positive user experience ratings (Mean Opinion Scores: 4.07 for naturalness, 3.98 for accent uniqueness, 3.88 for clarity).

Figure 2. Classification of artificial intelligence, machine learning, and algorithms in diabetes/prediabetes management. BERT representations from transformers. AI: artificial intelligence; BERT: Bidirectional Encoder; CNN: convolutional neural network; DNN: deep neural network; GPT-4: generative pre-trained transformer 4; KBOA: knowledge-based question answering; NLU: natural language understanding; RAG: retrieval-augmented generation; SVM: support vector machine; XGBoost: extreme gradient boosting [9,13,14,34-36,39-44,46,47,49,51,53-55,57-61,67,69,70].



Hybrid approaches such as the retriever-augmented generation model provided 98% accuracy in patient education for diabetes and diabetic foot care [39]. AI-driven interventions, such as the Sweetch app [67], used just-in-time adaptive interventions, improving adherence (82.9%) [8]. However, challenges included AI-specific barriers in 16% of studies, such as algorithmic bias [38,46], privacy concerns [59], and inconsistent responses [34], and alongside inadequate personalization (38% of studies) often due to generic content or lack of adaptive features [9,49,58]. Culturally tailored solutions, such as the DiaFriend app for Portuguese Americans [43] and AI-HEALS for Chinese patients [70], mitigated some barriers but faced limitations such as incomplete backends or language restrictions. These findings highlight AI's transformative potential in diabetes management while emphasizing the need for future large-scale, comparative, and longitudinal studies to determine their real-world effectiveness, cost-efficiency, and equity impact across diverse populations. Addressing biases, privacy, cultural adaptation, and sustained engagement challenges to optimize future implementations. Table 2 summarizes the comparison of AI-driven versus non-AI interventions.

XSL•FO
RenderX

Table . Comparison of artificial intelligence-driven versus non-artificial intelligence interventions.

Dimension	AI[a]	Partial-AI	Non-AI
Dropout range	6.6%-30% (n=3) [8,49,52].	6.4% (n=1) [55].	14.4%-35.5% (n=4) [53,57,61,64].
Engagement range (primary outcome)	63.9%-100% (n=4) [8,9,14,50].	Not mentioned.	65%-92% completion or retention (n=2) [48,53].
HbA^{b}_{1c} reduction range	0.2% - 1.8% (n=3) [9,50,52].	1% - 1.2% (n=2) [55,62].	0.19% - 1.6% (n=8) [13,37,47,53,56,58,61,63].
Barriers	Inadequate personalization [34,35,44,46,51], low engagement and adherence [34,44-46,50-52], technical issues and connectivity [35,44,51], short duration and follow-up [35,44,50-52], cultural and language issues [44], and poorly defined AI taxonomy [52].	Inadequate personalization [41,55,62], dropout and engagement issues [55,62], limited interactivity [55], cultural or language barriers [41,55], and limited clinician time [62].	Inadequate personalization [36,42,43,48,53,57,61,64], dropout or engagement issues [36,42,43,48,53,57,61,64], limited interactivity [42,43,48,53], and cultural or socioeconomic barriers [42,43,48,53,57,61]
Facilitators	Personalized feedback and adaptive algorithms [34,35,44,46,50-52], behavioral model integration [37,50,52], remote monitoring and provider feedback [44,50-52], user-friendly design [34,35,44,46,51], culturally tailored content [44], cloud integration for provider access [51,52], structured follow-up improving adherence [8,50,52], scalable interventions [39,50,52], integration of AI and human support [52], and emphasis on scalability and precision [52]	Tailored feedback and reminders [41,55,62], automated tracking features [41,55,62], behavioral frameworks (CBT[c], goal setting) [55,62], provider monitoring [55], educational and motivational support [41,55], cultural adaptation [41,55,62], and local food databases [55,62], real-time communication [55].	Educational content [42,43,48,53,57,61,64], human coaching and social support [42,43,48,53,58,61], culturally appropriate design [42,43,48,53,61], motivational reinforcement [48,53,58,61], health care professional guidance [42,48,53,58,61], simplicity and accessibility [48,53,58,61], integration with health care systems [42,48,53,58,61], supportive follow-up [48,53,58,61], structured educational design [13,63], emphasis on usability and accessibility [13,63], incorporation of behavioral science [13], positive patient–provider communication [13,63], and evidence synthesis improving generalizability [63].
Clinical outcomes	HbA_{1c} reduction [50,52], weight loss [50], improved self-management and adherence [44,50,52], improved patient satisfaction [35,44], no increase in hypoglycemia [52], scalability potential [39,50,52], positive usability outcomes [35,44], and clear trends toward improved glycemic control [52]	Improved glycemic control [41,55,62], weight loss [41,55], better engagement with mixed-mode support [55,62], improved knowledge and self-efficacy [55], feasibility demonstrated [55], and moderate-to-high user satisfaction [55].	Improved self-management [42,43,48,53,57,61,64], enhanced knowledge and motivation [42,43,48,53,57,61,64], weight reduction [48,53,61,64], HbA_{1c} reduction [48,53,61,64], improved patient confidence [42,43,48,53,61], high acceptability [58], positive behavioral outcomes [42,43]. Consistent improvement in self-care outcomes [13,63] reinforced the need for personalized interventions [13] and positive health literacy and behavioral impact [63].

[a]AI: artificial intelligence.

[b]HbA_{1c}: hemoglobin A1c.

[c]CBT: cognitive behavioral therapy.

Engagement Across Diverse Populations

Across the 37 [8,9,13,14,34-66] studies, only 8 explicitly targeted culturally or linguistically diverse populations, Portuguese Americans [43], Iranian adults [41,42], Singaporean users [45,55], Arab women [61], Hispanic men [36], and Chinese Americans [48] (Table 3). These studies collectively highlight how cultural adaptation enhances engagement and usability, though most lacked long-term quantitative evaluation. High engagement and retention were most evident in culturally grounded, community-based interventions. The Facebook-delivered DPP for Chinese Americans achieved 85% retention at one year [48], while the soccer-based Latino men's program retained 65% at 24 weeks [36]. Among Arabic-speaking women in Saudi Arabia, retention reached 100% despite cultural restrictions [61]. In contrast, prototype or design-phase studies in Iran and the US [41-43] did not measure engagement but emphasized usability and localized

XSL•FO
RenderX

content. AI-enabled programs from Singapore [45,55] demonstrated high usability (system usability scale 80.6) and low dropout (6.4%), reflecting benefits of linguistic personalization and local food databases in digitally literate populations. Common barriers included low digital literacy, gender or mobility restrictions, and limited multilingual functionality, while facilitators centered on language adaptation, cultural familiarity, and social support. Overall, cultural adaptation consistently improved acceptability, but few studies measured sustained engagement, an important focus for future research.

Table . Population diversity and engagement metrics across studies.

Population and study IDs	Country	AI[a] type	Cultural adaptation	Engagement or retention	Key findings	Main facilitators (F) and barriers (B)
Portuguese Americans [43]	USA	Non-AI	Portuguese food and visuals	Not measured	Prototype only and expected to improve adherence	Simple interface (F) and English-only backend (B)
Iranian [41,42]	Iran	AI or Partial-AI	Persian language, food DB[b], and TTM[c] tailoring	Usability only (short-term)	Positive clarity or usability and no outcome data	Localized design (F) and low digital literacy (B)
Singaporean [45,55]	Singapore	AI or Partial-AI	Multilingual TTS[d] and local food DB	SUS[e] 80.6, 6.4% dropout	High usability: engagement linked to weight loss	Personalization (F) and manual logging burden (B)
Arab women [68]	Saudi Arabia	Non-AI	Arabic-language and gender norms	100% retention	HbA^{f}_{1c} reduction ($P<.001$), feasible WhatsApp delivery	Cultural tailoring (F) and mobility limits (B)
Hispanic men [36]	USA	Non-AI	Bilingual coaches and soccer	65% retention	Improved fitness, motivation, and social bonding	Peer support (F) and time and work barriers (B)
Chinese Americans [48]	USA	Non-AI	Bilingual modules	85% retention (1y)	Improved satisfaction and 2.3% weight loss	Coach support (F) and low online literacy (B)

[a]AI: artificial intelligence.
[b]DB: dietary behavior.
[c]TTM: transtheoretical model.
[d]TTS: transtheoretical stage.
[e]SUS: system usability scale.
[f]HbA_{1c}: hemoglobin A1c.

Discussion

Overview

This systematic review synthesized engagement barriers, facilitators, and outcomes across AI-driven, partially AI, and non-AI digital health interventions for T2DM and prediabetes. By applying SDT and UCD as interpretive frameworks, this review extends prior work that has predominantly focused on clinical outcomes [27,48]. Our findings show that the most prevalent barriers to sustained user engagement were inadequate personalization, cultural or language mismatches, socioeconomic constraints, and, in AI tools, specific concerns about bias and privacy, while the strongest facilitators were personalized and adaptive feedback and cultural tailoring. These factors, through their influence on autonomy, competence, and relatedness and usability, appear to be the primary drivers of behavioral and clinical change in digital diabetes interventions.

Main Findings

Across the 3 7[8,9,13,14,34-66] included studies, digital health interventions demonstrated meaningful improvements in glycemic control, dietary behaviors, and physical activity, though sustained engagement remained a critical challenge. Inadequate personalization emerged as one of the most prevalent barriers, undermining SDT's principles of autonomy and competence and contributing to dropout. AI-based tools generally outperformed traditional digital programs when adaptivity, real-time monitoring, and tailored feedback were effectively implemented. For instance, AI-supported insulin titration and predictive analytics achieved faster glycemic improvements than standard care [5]. Interventions incorporating cultural tailoring, social support, and simplified interfaces consistently demonstrated higher engagement and completion rates. These findings emphasize that algorithmic sophistication alone is insufficient; meaningful engagement depends on how well digital systems address users' psychological needs, contextual realities, and cultural identities.

Engagement in Diverse Populations

Engagement varied substantially across demographic, cultural, and socioeconomic groups. Young adults frequently demonstrated low perceived risk and weaker intrinsic motivation to sustain engagement, while older adults faced usability challenges and digital literacy barriers. Cultural mismatch,

XSL•FO
RenderX

reported in nearly 40% of studies, led to reduced trust and relevance, particularly among minority groups.

Socioeconomic constraints such as limited smartphone access, high data costs, or inconsistent internet connectivity were especially apparent in LMIC settings. Interventions that provided multilingual content, culturally relevant dietary databases, or low-bandwidth delivery (eg, SMS text messaging-based chatbots) showed higher acceptability and engagement. These results emphasize the importance of context-aware and culturally grounded design practices when delivering digital diabetes interventions at scale.

Barriers to Engagement and Trust in AI

Several AI-specific barriers affected user trust and engagement. Algorithmic bias, arising from nonrepresentative training datasets, resulted in inaccurate risk predictions or poorly matched recommendations, particularly among ethnically diverse populations [42]. Privacy concerns were common, especially in cloud-based systems, with several users expressing discomfort about data security or opaque data handling processes [65]. Such concerns directly undermine SDT's relatedness and competence needs by diminishing the sense of transparency and credibility.

LLM-based or rule-based conversational agents occasionally produced generic, repetitive, or incorrect responses, which weakened trust and reduced perceived intervention quality. In contrast, systems that provided transparent rationales (eg, via Explainable AI), culturally adapted messaging, or adaptive learning mechanisms fostered significantly stronger engagement.

These insights highlight that strong technical performance does not guarantee user trust; trust must be actively cultivated through transparency, reliability, cultural sensitivity, and robust data governance.

Clinical and Behavioral Outcomes

Both AI-driven and non-AI digital interventions yielded positive clinical and behavioral outcomes. AI-enabled programs leveraging CGM, predictive modeling, or digital twins produced some of the largest HbA_{1c} reductions and behavioral improvements observed within the included studies [23,27]. Traditional digital interventions, such as structured online modules or SMS text messaging coaching, offered modest but consistent improvements in diet, physical activity, and self-management.

However, across all intervention types, the effectiveness of clinical outcomes was closely linked to user engagement. Interventions that successfully supported autonomy (eg, personalized goal setting), competence (eg, timely feedback), and relatedness (eg, social support) achieved higher adherence and more sustained improvements. These findings reinforce that engagement is not merely a process measure but a core determinant of intervention effectiveness.

Limitations

This review provides insights into digital health interventions for T2DM and prediabetes, but several limitations should be acknowledged. The Population, Intervention, Comparator, and Outcome-based search strategy may have missed studies using nonstandardized AI or digital health terms, and the English-only focus excluded relevant non-English or gray literature, potentially limiting the generalizability of findings to global populations. The review did not fully address informatics challenges, which are critical for scalability. Additionally, findings were not stratified by intervention type (AI-driven, partially AI-driven, vs non-AI) or by study setting and population due to limited quantitative data availability, which may restrict understanding of differences in engagement patterns and barriers across these subgroups.

The exclusion of studies targeting only children or older adults without broader adult data restricts insights into these populations, as older adults often have unique comorbidities and digital literacy challenges requiring separate evaluation. Methodological limitations of included studies, such as small sample sizes and heterogeneous designs, may affect generalizability. Quality appraisal was primarily conducted by a single reviewer and verified by a second, rather than through fully independent dual review, which may introduce subjective bias. Additionally, the inclusion of prior systematic reviews alongside primary studies introduces potential evidential overlap. We mitigated this by extracting only novel insights from the reviews; however, a small amount of overlap in the evidence is possible. This review provides a robust foundation for understanding engagement barriers and facilitators in digital health interventions for adults with T2DM and prediabetes.

Future Directions

The results from this review point to several important areas for future research. Longitudinal trials are needed to assess the long-term impacts of AI-driven interventions, particularly in terms of sustained engagement, clinical outcomes, and cost-effectiveness across diverse populations. Future studies should also explore AI-enabled solutions that integrate real-time data, such as CGM and wearables, to offer more precise, tailored interventions that enhance motivation and adherence.

Furthermore, inclusive research is needed to explore the effectiveness of digital health interventions in LMICs and underserved populations, where cultural sensitivity and accessibility are critical. Additionally, interoperability between AI tools and existing health care systems, such as electronic health records, must be addressed to facilitate seamless data sharing and personalized care. Research should focus on the integration of AI with established health care platforms, enabling a holistic approach to patient management.

Additionally, future studies should explore low-cost, offline solutions such as SMS text messaging-based multilingual chatbots, which can help bridge digital health access gaps in LMICs. Interdisciplinary collaborations between health informatics, behavioral science, and policy experts will be crucial for evaluating the scalability and ethics of these solutions globally. Ethical considerations, including data privacy, consent, and equitable access, must also be central to future research agendas.

XSL•FO
RenderX

Conclusions

This systematic review provides important insights into the design and implementation of digital health interventions for T2DM and prediabetes management, emphasizing the need for adaptive, inclusive, and user-centered solutions. Both AI-driven and non-AI interventions have shown promise in improving clinical outcomes and engagement, though each faces unique challenges. The integration of SDT and UCD principles, alongside advances in AI technology, can lead to more personalized and equitable solutions for diabetes care. Future research must prioritize diverse populations, cultural tailoring, and advanced informatics techniques to address current barriers and optimize the potential of digital health interventions in global diabetes prevention.

Funding

No external funding was provided for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy.
[DOC File, 18 KB - diabetes_v11i1e80582_app1.doc]

Multimedia Appendix 2
Detailed study characteristics and findings.
[DOC File, 47 KB - diabetes_v11i1e80582_app2.doc]

Checklist 1
PRISMA checklist.
[DOC File, 35 KB - diabetes_v11i1e80582_app3.doc]

References

1. Diabetes facts and figures. International Diabetes Federation. URL: https://www.idf.org [accessed 2026-01-27]
2. Sun H, Saeedi P, Karuranga S, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
3. Godman B, Basu D, Pillay Y, et al. Review of ongoing activities and challenges to improve the care of patients with type 2 diabetes across africa and the implications for the future. Front Pharmacol 2020;11:108. [doi: 10.3389/fphar.2020.00108] [Medline: 32265688]
4. Sami W, Ansari T, Butt NS, Hamid MRA. Effect of diet on type 2 diabetes mellitus: a review. Int J Health Sci (Qassim) 2017;11(2):65-71. [Medline: 28539866]
5. Reed J, Bain S, Kanamarlapudi V. A review of current trends with type 2 diabetes epidemiology, aetiology, pathogenesis, treatments and future perspectives. Diabetes Metab Syndr Obes 2021;14:3567-3602. [doi: 10.2147/DMSO.S319895] [Medline: 34413662]
6. Sui CF, Ming LC, Soh YC, Ng CH, Al-Worafi YM, Hussain Z. Systematic review protocol for effectiveness and cost-effectiveness of non-surgical interventions to prevent diabetes progression in adults with prediabetes. BMJ Open 2025 Jul 22;15(7):e093768. [doi: 10.1136/bmjopen-2024-093768] [Medline: 40701593]
7. Gardner DSL, Saboo B, Kesavadev J, et al. Digital health technology in diabetes management in the Asia-Pacific region: a narrative review of the current scenario and future outlook. Diabetes Ther 2025 Mar;16(3):349-369. [doi: 10.1007/s13300-025-01692-0] [Medline: 39928223]
8. Nayak A, Vakili S, Nayak K, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023 Dec 1;6(12):e2340232. [doi: 10.1001/jamanetworkopen.2023.40232] [Medline: 38039007]
9. Wu Y, Zhang J, Ge P, et al. Application of chatbots to help patients self-manage diabetes: systematic review and meta-analysis. J Med Internet Res 2024 Dec 3;26(10397):e60380. [doi: 10.2196/60380] [Medline: 39626235]
10. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
11. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]

XSL•FO
RenderX

12. Mbanugo OJ. AI-enhanced telemedicine: a common-sense approach to chronic disease management and a tool to bridging the gap in healthcare disparities. Int J Res Publ Rev 2025;6(2):3223-3241. [doi: 10.55248/gengpi.6.0225.0952]
13. Xue H, Zhang L, Shi Y, et al. The effectiveness of digital health intervention on glycemic control and physical activity in patients with type 2 diabetes: a systematic review and meta-analysis. Front Digit Health 2025;7(1630588):1630588. [doi: 10.3389/fdgth.2025.1630588] [Medline: 40799285]
14. Oh YJ, Zhang J, Fang ML, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160. [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
15. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68] [Medline: 11392867]
16. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction: Taylor & Francis group; 1986. [doi: 10.1201/b15703]
17. Prince G, Rees Lewis D, Pollack T, et al. Employing user-centered design and education sciences to inform training of diabetes survival skills. J Clin Transl Endocrinol 2024 Sep;37:100364. [doi: 10.1016/j.jcte.2024.100364] [Medline: 39247534]
18. Mathiesen AS, Zoffmann V, Lindschou J, et al. Self-determination theory interventions versus usual care in people with diabetes: a systematic review with meta-analysis and trial sequential analysis. Syst Rev 2023 Sep 6;12(1):158. [doi: 10.1186/s13643-023-02308-z] [Medline: 37674180]
19. Cochrane Handbook for Systematic Reviews of Interventions, th edition: Wiley; 2019. [doi: 10.1002/9781119536604]
20. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
21. Optimizing engagement in digital health interventions for diabetes and prediabetes: a systematic review. OSF. URL: https://doi.org/10.17605/OSF.IO/FC752 [accessed 2026-01-27]
22. Razavi S. Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. Environ Model Softw 2021 Oct;144:105159. [doi: 10.1016/j.envsoft.2021.105159]
23. Bahadori S, Buckle P, Soukup Ascensao T, Ghafur S, Kierkegaard P. Evolving digital health technologies: aligning with and enhancing the national institute for health and care excellence evidence standards framework. JMIR Mhealth Uhealth 2025 Aug 22;13:e67435. [doi: 10.2196/67435] [Medline: 40845677]
24. Carpenter R, DiChiacchio T, Barker K. Interventions for self-management of type 2 diabetes: an integrative review. Int J Nurs Sci 2019 Jan 10;6(1):70-91. [doi: 10.1016/j.ijnss.2018.12.002] [Medline: 31406872]
25. Benton JS, Cotterill S, Hawkes RE, Miles LM, French DP. Changes in a digital type 2 diabetes self-management intervention during national rollout: mixed methods study of fidelity. J Med Internet Res 2022 Dec 7;24(12):e39483. [doi: 10.2196/39483] [Medline: 36476723]
26. Glechner A, Keuchel L, Affengruber L, et al. Effects of lifestyle changes on adults with prediabetes: a systematic review and meta-analysis. Prim Care Diabetes 2018 Oct;12(5):393-408. [doi: 10.1016/j.pcd.2018.07.003] [Medline: 30076075]
27. Wang Y, Chai X, Wang Y, et al. Effectiveness of different intervention modes in lifestyle intervention for the prevention of type 2 diabetes and the reversion to normoglycemia in adults with prediabetes: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025 Jan 29;27:e63975. [doi: 10.2196/63975] [Medline: 39879607]
28. Higgins JPT, Altman DG, Gøtzsche PC, et al. The cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
29. Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. Research Methods in Medicine & Health Sciences 2020 Sep;1(1):31-42. [doi: 10.1177/2632084320947559]
30. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008. [doi: 10.1136/bmj.j4008] [Medline: 28935701]
31. Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). J Eval Clin Pract 2018 Jun;24(3):459-467. [doi: 10.1111/jep.12884] [Medline: 29464873]
32. Pluye P, Gagnon MP, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
33. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: effectiveness of interventions to promote smoke alarm ownership and function. Evaluation (Lond) 1995;15(1):49-73. [doi: 10.1177/1356389008097871]
34. Hussain W, Grundy J. Advice for diabetes self-management by chatgpt models: challenges and recommendations. Arxiv. Preprint posted online on 2025. [doi: 10.48550/arxiv.2501.07931] [Medline: 41333165]
35. Sun H, Zhang K, Lan W, et al. An AI dietitian for type 2 diabetes mellitus management based on large language and image recognition models: preclinical concept validation study. J Med Internet Res 2023 Nov 9;25(2):e51300. [doi: 10.2196/51300] [Medline: 37943581]

XSL•FO
RenderX

36. Frediani JK, Bienvenida AF, Li J, Higgins MK, Lobelo F. Physical fitness and activity changes after a 24-week soccer-based adaptation of the U.S diabetes prevention program intervention in Hispanic men. Prog Cardiovasc Dis 2020;63(6):775-785. [doi: 10.1016/j.pcad.2020.06.012] [Medline: 32603753]
37. Ruiz-Leon AM, Casas R, Castro-Barquero S, et al. Efficacy of a mobile health-based behavioral treatment for lifestyle modification in type 2 diabetes self-management: greenhabit randomized controlled trial. J Med Internet Res 2025 Jan 22;27:e58319. [doi: 10.2196/58319] [Medline: 39841995]
38. Wang SCY, Nickel G, Venkatesh KP, Raza MM, Kvedar JC. AI-based diabetes care: risk prediction models and implementation concerns. NPJ Digit Med 2024 Feb 15;7(1):36. [doi: 10.1038/s41746-024-01034-7] [Medline: 38361152]
39. Mashatian S, Armstrong DG, Ritter A, et al. Building trustworthy generative artificial intelligence for diabetes care and limb preservation: a medical knowledge extraction case. J Diabetes Sci Technol 2025 Sep;19(5):1264-1270. [doi: 10.1177/19322968241253568] [Medline: 38767382]
40. Alloatti F, Bosca A, Di Caro L, Pieraccini F. Diabetes and conversational agents: the AIDA project case study. Discov Artif Intell 2021;1(1):4. [doi: 10.1007/s44163-021-00005-1] [Medline: 40478072]
41. Salari R, R Niakan Kalhori S, GhaziSaeedi M, Jeddi M, Nazari M, Fatehi F. Mobile-Based and cloud-based system for self-management of people with type 2 diabetes: development and usability evaluation. J Med Internet Res 2021 Jun 2;23(6):e18167. [doi: 10.2196/18167] [Medline: 34076579]
42. Mehraeen E, Mehrtak M, Janfaza N, et al. Design and development of a mobile-based self-care application for patients with type 2 diabetes. J Diabetes Sci Technol 2022 Jul;16(4):1008-1015. [doi: 10.1177/19322968211007124] [Medline: 33840235]
43. LeSeure P, Chin E, Zhang S. A culturally sensitive mobile app (DiaFriend) to improve self-care in patients with type 2 diabetes: development study. JMIR Diabetes 2024 Oct 21;9:e63393. [doi: 10.2196/63393] [Medline: 39432893]
44. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
45. Nguyen TT, Sim K, Anthony To Yiu K, O'Donnell RR, Lim ST, Wang W, et al. arXiv. Designing AI-Based Conversational Agent for Diabetes Care in a Multilingual Context. Preprint posted online on May 20, 2021. [doi: 10.48550/arXiv.2105.09490]
46. Shiferaw MW, Zheng T, Winter A, Mike LA, Chan LN. Assessing the accuracy and quality of artificial intelligence (AI) chatbot-generated responses in making patient-specific drug-therapy and healthcare-related decisions. BMC Med Inform Decis Mak 2024 Dec 24;24(1):404. [doi: 10.1186/s12911-024-02824-5] [Medline: 39719573]
47. de Souza Ferreira E, de Aguiar Franco F, Dos Santos Lara MM, et al. The effectiveness of mobile application for monitoring diabetes mellitus and hypertension in the adult and elderly population: systematic review and meta-analysis. BMC Health Serv Res 2023 Aug 12;23(1):855. [doi: 10.1186/s12913-023-09879-6] [Medline: 37573312]
48. Yeh MC, Lau W, Keady CA, et al. Evaluation of feasibility and acceptability of a web-based diabetes prevention program (DPP) for diabetes risk reduction in Chinese Americans in New York City. Front Public Health 2023;11:1199746. [doi: 10.3389/fpubh.2023.1199746] [Medline: 37333528]
49. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845. [doi: 10.2196/22845] [Medline: 32996892]
50. Mathioudakis N, Lalani B, Abusamaan MS, et al. An AI-powered lifestyle intervention vs human coaching in the diabetes prevention program. JAMA 2025 Dec 16;334(23):2079-2089. [doi: 10.1001/jama.2025.19563] [Medline: 41144242]
51. Elhadd T, Mall R, Bashir M, et al. Artificial Intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during ramadan (The PROFAST - IT Ramadan study). Diabetes Res Clin Pract 2020 Nov;169:108388. [doi: 10.1016/j.diabres.2020.108388] [Medline: 32858096]
52. Shamanna P, Erukulapati RS, Shukla A, et al. One-year outcomes of a digital twin intervention for type 2 diabetes: a retrospective real-world study. Sci Rep 2024 Oct 26;14(1):25478. [doi: 10.1038/s41598-024-76584-7] [Medline: 39461977]
53. Bailey-Davis L, Wood GC, Cook A, et al. Communicating personalized risk of diabetes and offering weight reduction program choice: recruitment, participation, and outcomes. Patient Educ Couns 2021 May;104(5):1193-1199. [doi: 10.1016/j.pec.2020.10.017] [Medline: 33097360]
54. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
55. Lim SL, Tay MHJ, Ong KW, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022 Sep 30;7(3):e35039. [doi: 10.2196/35039] [Medline: 36178718]
56. Li M, Liu S, Yu B, et al. Assessing the effectiveness of digital health behavior strategies on type 2 diabetes management: systematic review and network meta-analysis. J Med Internet Res 2025 Feb 14;27:e63209. [doi: 10.2196/63209] [Medline: 39951722]

XSL•FO
RenderX

57. Nanditha A, Thomson H, Susairaj P, et al. A pragmatic and scalable strategy using mobile technology to promote sustained lifestyle changes to prevent type 2 diabetes in India and the UK: a randomised controlled trial. Diabetologia 2020 Mar;63(3):486-496. [doi: 10.1007/s00125-019-05061-y] [Medline: 31919539]
58. Green JB, Crowley MJ, Thirunavukkarasu S, Maruthur NM, Oldenburg B. The final frontier in diabetes care: implementing research in real-world practice. Diabetes Care 2024 Aug 1;47(8):1299-1310. [doi: 10.2337/dci24-0001] [Medline: 38907682]
59. Jabara M, Kose O, Perlman G, et al. Artificial intelligence-based digital biomarkers for type 2 diabetes: a review. Can J Cardiol 2024 Oct;40(10):1922-1933. [doi: 10.1016/j.cjca.2024.07.028] [Medline: 39111729]
60. Nadarzynski T, Knights N, Husbands D, et al. Achieving health equity through conversational AI: a roadmap for design and implementation of inclusive chatbots in healthcare. PLOS Digit Health 2024 May;3(5):e0000492. [doi: 10.1371/journal.pdig.0000492] [Medline: 38696359]
61. Al-Hamdan R, Avery A, Al-Disi D, Sabico S, Al-Daghri NM, McCullough F. Efficacy of lifestyle intervention program for Arab women with prediabetes using social media as an alternative platform of delivery. J Diabetes Investig 2021 Oct;12(10):1872-1880. [doi: 10.1111/jdi.13531] [Medline: 33638248]
62. Bergenstal RM, Johnson M, Passi R, et al. Automated insulin dosing guidance to optimise insulin management in patients with type 2 diabetes: a multicentre, randomised controlled trial. Lancet 2019 Mar 16;393(10176):1138-1148. [doi: 10.1016/S0140-6736(19)30368-X] [Medline: 30808512]
63. Xiao Y, Wang Z, Zhang L, et al. Effectiveness of digital diabetes management technology on blood glucose in patients with type 2 diabetes at home: systematic review and meta-analysis. J Med Internet Res 2025 Mar 3;27:e66441. [doi: 10.2196/66441] [Medline: 40053775]
64. Röhling M, Kempf K, Banzer W, et al. Prediabetes conversion to normoglycemia Is superior adding a low-carbohydrate and energy deficit formula diet to lifestyle intervention-a 12-month subanalysis of the ACOORH trial. Nutrients 2020 Jul 7;12(7):2022. [doi: 10.3390/nu12072022] [Medline: 32646010]
65. Nomura A, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial intelligence in current diabetes management and prediction. Curr Diab Rep 2021 Dec 13;21(12):61. [doi: 10.1007/s11892-021-01423-2] [Medline: 34902070]
66. Lim D, Meier L, Cadwell KM, Jacob C. From diabetes care to prevention: review of prediabetes apps in the DACH region. Mhealth 2025;11:8. [doi: 10.21037/mhealth-24-57] [Medline: 39944857]
67. Abusamaan MS, Ballreich J, Dobs A, et al. Effectiveness of artificial intelligence vs. human coaching in diabetes prevention: a study protocol for a randomized controlled trial. Trials 2024 May 16;25(1):325. [doi: 10.1186/s13063-024-08177-8] [Medline: 38755706]
68. Al‐Hamdan R, Avery A, Al‐Disi D, Sabico S, Al‐Daghri NM, McCullough F. Efficacy of lifestyle intervention program for Arab women with prediabetes using social media as an alternative platform of delivery. J of Diabetes Invest 2021 Oct;12(10):1872-1880 [FREE Full text] [doi: 10.1111/jdi.13531]
69. Röhling M, Kempf K, Banzer W, et al. Prediabetes conversion to normoglycemia Is superior adding a low-carbohydrate and energy deficit formula diet to lifestyle intervention-A 12-month subanalysis of the ACOORH trial. Nutrients 2020 Jul 7;12(7):2022. [doi: 10.3390/nu12072022] [Medline: 32646010]
70. Wu Y, Min H, Li M, et al. Effect of artificial intelligence-based health education accurately linking system (AI-HEALS) for type 2 diabetes self-management: protocol for a mixed-methods study. BMC Public Health 2023;23(1). [doi: 10.1186/s12889-023-16066-z]

Abbreviations

AI: artificial intelligence
CGM: continuous glucose monitors
DPP: diabetes prevention program
LMIC: low-middle income countries
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SDT: self-determination theory
T2DM: type 2 diabetes mellitus
UCD: user-centered design

XSL•FO
RenderX

Edited by L Quinlan; submitted 13.Jul.2025; peer-reviewed by S Bolen, T Duong; revised version received 29.Nov.2025; accepted 19.Dec.2025; published 12.Mar.2026.

Please cite as:
Thanthrige A, Wickramasinghe N
Digital Health Solutions for Type 2 Diabetes and Prediabetes: Systematic Review of Engagement Barriers, Facilitators, and Outcomes
JMIR Diabetes 2026;11:e80582
URL: https://diabetes.jmir.org/2026/1/e80582
doi:10.2196/80582

© Ayesha Thanthrige, Nilmini Wickramasinghe. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 12.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness, Reach, Uptake, and Feasibility of Digital Health Interventions for Culturally and Linguistically Diverse Populations Living With Prediabetes Across the Lifespan: Systematic Review and Meta-Analysis

Lisa Whitehead[1,2,3,4], BSc (Hons), MA, PhD; Min Zhang[1,4], MSCN; Wai Hang Kwok[1], PhD; Diana Arabiat[1,2,4,5], PhD; Amanda Towell Barnard[1], PhD

[1]School of Nursing and Midwifery, Edith Cowan University, 270 Joondalup Drive, Joondalup, Australia

[2]JBI – The University of Jordan Health Research Centre, The University of Jordan, Amman, Jordan

[3]The Department of Nursing, University of Otago, Christchurch, New Zealand

[4]JBI Centre for Evidence Informed Nursing and Midwifery, Edith Cowan University, Joondalup, Australia

[5]Epidemiology, Biostatistics and Health Informatics Department, Public Health Institute (PHI), The University of Jordan, Amman, Jordan

Corresponding Author:
Lisa Whitehead, BSc (Hons), MA, PhD
School of Nursing and Midwifery, Edith Cowan University, 270 Joondalup Drive, Joondalup, Australia

Abstract

Background: Culturally and linguistically diverse (CaLD) populations are at a higher risk of developing prediabetes; however, the effectiveness and implementation of digital health interventions for prediabetes management in this population are not well understood.

Objective: This review aims to evaluate the effectiveness and implementation of digital health interventions (DHIs) versus usual care for glycemic control in CaLD populations living with prediabetes.

Methods: This review aimed to include people of any age living with prediabetes who are from a CaLD background. Experimental and quasi-experimental studies that compare digital health interventions to usual care, waitlist, or active control were eligible. The primary outcome was glycemic control as measured by hemoglobin A_{1c}. A comprehensive search was conducted in CINAHL, Cochrane Library, Embase, MEDLINE, 3 trial registers, and gray literature databases, along with reference lists for additional studies. Studies published in English and published since the inception of each database were included. Statistical analyses included meta-analysis, sensitivity analyses, subgroup analyses, meta-regression, and publication bias assessments. The methodological quality was assessed using the JBI critical appraisal tools, and the quality of evidence was evaluated using Grading of Recommendations, Assessment, Development, and Evaluation to create summary of findings tables. Random-effects models with restricted maximum likelihood estimation were employed.

Results: A total of 14 studies involving 5714 adult participants were included. The meta-analysis showed that DHIs were associated with a reduction in hemoglobin A_{1c} (P<.001), though evidence certainty was low (mean difference=−0.14, 95% CI −0.24 to −0.05). Effects on fasting blood glucose and body weight remain uncertain. Implementation outcomes demonstrated high uptake (>78.8%), engagement (>80%), and intention rates (89.1%) among CaLD populations with prediabetes. Significant heterogeneity was observed in both randomized controlled trials and pre-post studies. Subgroup analyses revealed significant effects at the 6-month follow-up point only for interventions (P<.001). Meta-regression identified comorbidity status as the only significant contributor to heterogeneity (P=.02). Sensitivity analyses demonstrated robust significant effects (P<.001). Publication bias assessment showed mixed results (Begg P=.23, Egger P=.02), but trim-and-fill analysis confirmed the robustness of the findings with no missing studies. Despite these positive findings, substantial heterogeneity across most outcomes and low-to-very low certainty evidence limit the reliability of these results, warranting cautious interpretation.

Conclusions: DHIs demonstrate potential for improving glycemic control in CaLD populations living with prediabetes. The observed heterogeneity could be attributed to intervention duration, control type, and participants' comorbidity status. While the findings related to implementation were encouraging, the certainty of the evidence and substantial heterogeneity suggest that DHIs should be used as adjunctive tools with health care provider involvement rather than stand-alone solutions due to low certainty evidence and substantial heterogeneity. Further rigorous research considering contextual, individual, and cultural factors is needed.

Trial Registration: PROSPERO CRD42024556292; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024556292

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e70912) doi:10.2196/70912

KEYWORDS

culturally and linguistically diverse; digital health; glycemic control; meta-analysis; prediabetic state; systematic review

Introduction

Background

The global burden of prediabetes is substantial and increasing. Prediabetes, also known as impaired glucose tolerance (IGT) and impaired fasting glucose (IFG), increases the risk of developing type 2 diabetes, cardiovascular disease, and stroke [1,2]. Once diagnosed with prediabetes and without intervention, 5% - 10% of the people per year will progress to a diagnosis of type 2 diabetes, and this rate is higher for specific population groups (eg, South and East Asian people and older adults) [3]. Worldwide, the prevalence of IGT and IFG was 9.1% (464 million) and 5.8% (298 million) in 2021, respectively [2]. High-income countries had the highest rates of prediabetes in 2021, and low-income countries are projected to experience a significant rise in prevalence by 2045 [2]. For some individuals living with prediabetes, early and timely treatment, such as lifestyle changes, can effectively prevent or delay the onset of type 2 diabetes [4]. Thus, effective intervention strategies are needed in low-, middle-, and high-income countries to address the diabetes epidemic.

The use of digital health technologies is well established in diabetes management to support patient self-care [5-7]. Digital health, or eHealth, includes a range of approaches, such as mobile health, telehealth and telemedicine, health information technology, wearable devices, and personalized medicine [8]. Digital health interventions (DHIs) are programs that provide information, communication, support, and networks to people to improve their physical and mental health through the use of digital technologies such as smartphones, websites, and text messaging [7,9]. DHIs can help facilitate tailored interventions and improve accessibility for hard-to-reach populations to affect behavior change [10].

Previous reviews have suggested that DHIs, such as digital health coaching, technology-assisted diabetes prevention programs, and digital health–supported lifestyle change programs, are promising strategies to support people living with prediabetes [11-15], type 1 and type 2 diabetes, and their cardiovascular complications [12-19]. Although there is growing evidence to support the use of DHIs among adults with diabetes or prediabetes, further research is needed to establish the efficacy of DHIs in improving prediabetes-related outcomes among diverse groups [11,15,20]. For example, a recent scoping review reported that while DHIs are acceptable for prediabetes self-management, their effectiveness in reducing the risk of type 2 diabetes is still inconclusive [20].

Studies have reported that people from culturally and linguistically diverse (CaLD) backgrounds have a higher risk of diabetes, associated hospitalizations, and mortality but lower uptake and use of DHIs compared to non-CaLD groups [9,21-23]. The term CaLD in this review is defined as "people born in non-English-speaking countries and/or who do not speak English at home" [23]. This category includes racial or ethnic minority groups, immigrants, and refugees [9]. To promote health equity among CaLD groups living with prediabetes, a comprehensive understanding of the effectiveness, reach, uptake, and feasibility of DHIs within CaLD populations is needed.

A systematic search of PROSPERO, MEDLINE, the Cochrane Database of Systematic Reviews, and the Joanna Briggs Institute (JBI) Evidence Synthesis was conducted, and no ongoing or planned systematic reviews on this targeted population were identified. To address this gap, this systematic review aims to synthesize the current evidence on the effectiveness, reach, uptake, and feasibility of DHIs among CaLD living with prediabetes across the lifespan. The outcome reporting is informed by the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework, which offers a systematic approach to evaluate the implementation and translational potential of health interventions [24,25].

Review Questions

The review questions were as follows: (1) What is the effectiveness of DHIs versus usual care, waitlist, or active comparator on glycemic control in CaLD populations living with prediabetes? (2) How do these interventions compare in terms of their reach, uptake, and feasibility in CaLD populations living with prediabetes?

Methods

Overview

This proposed systematic review was conducted following the JBI methodology for systematic reviews of effectiveness and the *Cochrane Handbook for Systematic Reviews of Interventions* to ensure rigorous standards in addressing methodological issues in meta-analyses. We also used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist to ensure clear reporting (Checklist 1) [26]. The protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42024556292).

Inclusion Criteria

Participants

Studies that included people of any age living with prediabetes from a CaLD background were considered for inclusion in this review without any limitations based on their gender, diagnostic criteria, or duration of the disease. Prediabetes is defined by the presence of IFG, or IGT, or an elevated hemoglobin A_{1c} (HbA_{1c}).

We used an evidence-based definition of CaLD study participants, which refers to those who were born in non-English–speaking countries or those whose main language is not English (eg, immigrants and refugees). Indigenous peoples

XSL•FO
RenderX

(eg, First Nations and Indigenous peoples in Australia, Canada, the United States, and New Zealand) were not included in this review as they are classified under a separate definition [23]. Thus, this review included CaLD groups that were (1) born in countries where the official language differs from that of their current country of residence or whose language spoken at home is not the official language of the country where they reside or (2) populations that were described in studies as "ethnically or racially diverse" or "ethnic or racial minority" [9].

Interventions

This review considered studies evaluating DHIs. A DHI is defined as a discrete technology functionality or capability designed to achieve a specific objective addressing a health system challenge [27]. We categorized the DHIs into 2 groups: the targeted primary user and the stand-alone or integrated interventions. According to the "Classification of Digital Interventions, Services, and Applications in Health" proposed by the World Health Organization, the targeted primary user category included participants or caregivers, health care providers, managers, and data services [27]. The category of stand-alone or integrated interventions includes independent digital components, such as SMS text messaging services, smartphones or tablets, websites, computer-based programs, and videoconferencing. It also encompasses integrated interventions that combine different digital components. There was no limitation in relation to the intensity, frequency, or duration of interventions.

Comparators

This review considered studies that compared DHIs to usual care, waitlist, or an active control group. An active control was defined as a specially designed intervention (eg, physical activity, diet counseling, and health education) delivered either face-to-face or through printed materials.

Outcomes

This review considered studies that included the outcomes of effectiveness, reach, uptake, or feasibility of DHIs.

Evaluation of Intervention Effectiveness

Effectiveness was defined as "the impact of an intervention on important outcomes, including potential negative effects, quality of life, and economic outcomes" [28]. The primary outcome in this review was the impact on HbA1c (%). Secondary outcomes included fasting plasma glucose (FPG), anthropometric indices (eg, body mass index, weight, waist circumference), and patient-reported outcomes.

Evaluation of Implementation Outcomes

Implementation outcomes included reach, uptake, engagement, and feasibility based on the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework. Reach was defined as "the absolute number, proportion, and representativeness of individuals who are willing to participate in a given initiative, intervention, or program" [28]. In this review, we operationalized intervention reach using the eligibility percentage, defined as: Intervention Reach=(Number of eligible participants/Total number of individuals invited)×100%.

Uptake (or adoption) was operationalized as the action of taking up or using the intervention or health promotion program components [28]. We considered reach and uptake at the individual participant level. Intervention uptake was captured through the uptake rate, calculated as follows: Uptake Rate=[(Number of individuals who started participating in the intervention/Total number of eligible individuals invited)×100%], as well as through narrative descriptions of participation (eg, the degree of participation in various components of the intervention).

For feasibility, we considered all information on participant satisfaction, acceptance, adherence, retention rates, and user feedback to gain a nuanced understanding of how interventions were received and implemented.

Types of Studies

Following the Cochrane "Algorithm to decide whether a review should include non-randomized studies of an intervention or not" [29], we included both experimental and quasi-experimental studies, including randomized controlled trials (pilot RCTs, crossover RCTs, cluster RCTs, and prospective RCTs), nonrandomized controlled trials, pre-post studies, and interrupted time-series studies.

Search Strategy

The search strategy aimed to locate peer-reviewed published studies. This review utilized a 3-step search strategy. An initial search of MEDLINE (EBSCO) was conducted to identify relevant articles on the topic. A full search strategy, including all identified keywords and index terms, was developed with the assistance of a research librarian.

A comprehensive search was then conducted in 4 databases: CINAHL, Cochrane Library, Embase, and MEDLINE; 3 trial register websites, including ClinicalTrials.gov, the Australian New Zealand Clinical Trials Registry, and the International Clinical Trials Registry Platform; as well as 2 gray literature websites, including ProQuest Dissertations & Theses and OpenGrey.EU. The search strategy was adjusted to suit each database and website included in this review (see the full search strategy in Table S1 in Multimedia Appendix 1).

The reference lists of systematic reviews on similar topics were screened for additional studies. Studies published in English since the inception of each database were included.

Study Selection

Following the search, all identified citations were collated and uploaded into EndNote (version 20.0; Clarivate Analytics), and duplicates were removed. Following a pilot test, titles and abstracts were screened by 2 independent reviewers for assessment against the inclusion criteria for the review (LW, MZ, WHK). Potentially relevant studies were retrieved in full, and their citation details were imported into the JBI System for the Unified Management, Assessment, and Review of Information (JBI SUMARI) (JBI) [26].

Multiple independent reviewers performed the full-text screening for each record according to the inclusion criteria (ATB, LW, MZ, and WHK). The reasons for the exclusion of

XSL•FO
RenderX

full-text studies were recorded. Any disagreements that arose between the reviewers at each stage of the selection process were resolved through discussion with a third reviewer (LW).

Assessment of Methodological Quality

The revised JBI critical appraisal tool for the assessment of risk of bias for RCTs [30] and quasi-experimental studies [31] was used by 2 independent reviewers (LW and MZ) to assess the methodological quality of the included studies. Any disagreements that arose were resolved through discussion with a third reviewer (ATB). Critical appraisal results were reported in a narrative form and presented in a table. All studies, regardless of their methodological quality results, underwent data extraction and synthesis.

Data Extraction

Two independent reviewers (LW and MZ) used a standardized JBI data extraction tool in JBI SUMARI (JBI) to extract data. The data extracted included information regarding the participants' characteristics (eg, age, sample size, percentage of female participants, ethnicity), study methods (eg, study design, country, and settings), intervention details (eg, duration, frequency, components, mode of delivery), and outcomes of significance to the review objective (primary and secondary outcomes). Any disagreements that arose during data extraction were resolved by the decision of a third reviewer (DA).

The authors of the studies were contacted to request missing data or clarification, where required. Where a study was reported in multiple publications, the earliest publication was included, and subsequent ones were identified as duplicates.

Data Synthesis

Data synthesis was conducted through a statistical meta-analysis and narrative synthesis. Statistical analyses were performed using JBI SUMARI and Stata (version 17.0; JBI) by pooling data from the included studies and generating forest plots, funnel plots, and bubble plots. The final postintervention mean differences were used to present the effect size for continuous data (eg, HbA_{1c}, weight). Forest plots present mean differences to account for heterogeneity, with corresponding 95% CIs.

The heterogeneity between the studies was evaluated using the I^2 (1% - 100%) statistic and visualized with Galbraith plots. Due to substantial heterogeneity (I^2>75%) and the inclusion of pre-post studies, random-effects meta-analyses using the restricted maximum likelihood method were used to pool the data. The sources of heterogeneity were explored through meta-regression, sensitivity analyses, and subgroup analyses. Sensitivity analyses were performed using a leave-one-out meta-analysis to test decisions made regarding the effectiveness of interventions versus comparators on HbA_{1c} levels. Subgroup analyses were conducted based on intervention characteristics and types of control. A qualitative evaluation of the heterogeneity of the included studies was also conducted by comparing participant and study characteristics.

Given that our meta-analysis included 10 or more studies, we performed funnel plot asymmetry analysis, including the Egger test and Begg test, to assess publication bias. Although no significant publication bias was detected, given the substantial heterogeneity (I^2>50%), a trim-and-fill test was performed to assess the robustness of the pooled results and estimate the potential impact of missing studies. Secondary outcomes were meta-analyzed when sufficient data were available. Furthermore, implementation outcomes were narratively synthesized and presented with supporting tables and figures due to the inappropriateness of statistical pooling.

Assessing Certainty in the Findings

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach was followed to grade the certainty of evidence [32]. A Summary of Findings (SoF) table was generated using the web-based software GRADEpro GDT/2015 (McMaster University, ON, Canada) to summarize the strength and reliability of the evidence. At least 2 independent reviewers (ATB and MZ) initially undertook this at the primary outcome level. Any disagreements that arose between the reviewers were resolved through discussion with a third reviewer (LW). Where required, the authors of included studies were contacted to request missing or additional data for clarification. The SoF table presents a ranking of the quality of the evidence based on the risk of bias, indirectness, inconsistency, imprecision, and publication bias. The outcomes reported in the SoF table were HbA_{1c} (%), FPG, and body weight.

Results

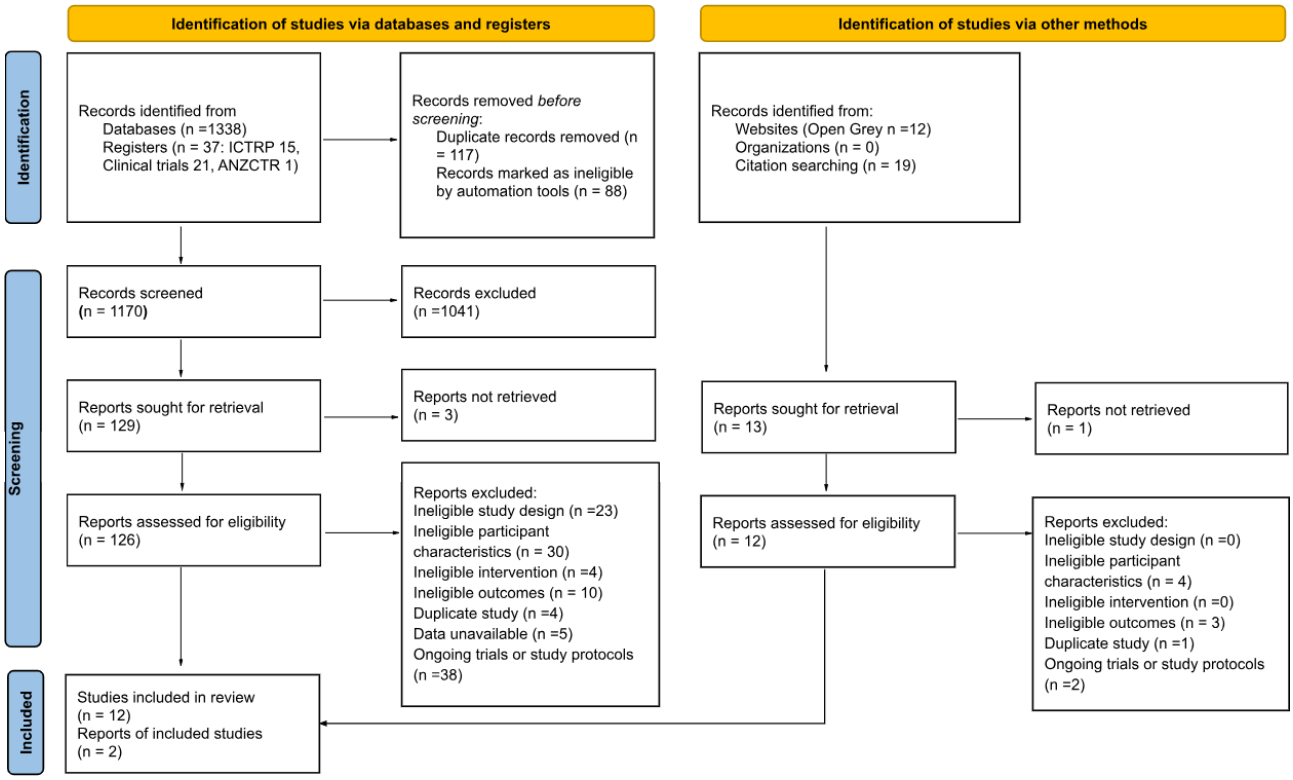
Study Inclusion

Our initial search yielded 1406 records. After removing 205 duplicate records and those marked as ineligible using EndNote, the remaining articles were imported into JBI SUMARI for further screening, as shown in the PRISMA flow diagram. We removed 1041 articles by assessing titles and abstracts, and a total of 138 articles (126 from databases and trial registries and 12 from website and citation searching) were retrieved for full-text review.

The search results and the study selection process were reported according to a PRISMA flow diagram in the final review (Figure 1) [33]. Based on the exclusion criteria shown in Figure 1, a total of 14 articles were eligible for inclusion in the systematic review. However, as 1 multinational study reported separate data for each region and another study reported outcomes for 2 distinct intervention completion levels, 17 independent datasets were included in the meta-analysis. The list of excluded studies and their reasons for exclusion is provided in Table S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. ANZCTR: Australian New Zealand Clinical Trials Registry; ICTRP: International Clinical Trials Registry Platform.



Methodological Quality

A total of 14 studies underwent methodological quality assessment. Nine RCTs demonstrated high methodological quality, reporting "Yes" for at least 9 of 13 quality assessment items (Table S3 in Multimedia Appendix 1). All RCTs used intention-to-treat analysis, meaning they included participants in the groups to which they were originally assigned. However, all studies either had unclear results or reported no blinding for participants (0%) and treatment providers (0%). Only 22.22% (2 studies) had outcome assessors who were blind to the treatment assignment. According to Khunti et al. [34], this lack of blinding may be due to the nature of DHIs.

In addition, 5 pre-post studies showed moderate quality, with ≤4 of 9 items rated as unclear or not met (Table S3 in Multimedia Appendix 1). As single-group studies, they inherently could not meet 2 specific items: "There was a control group" (rated as "No") and "Participants receiving similar treatment/care other than the intervention of interest"(rated as "N/A" due to the lack of between-group comparisons).

Characteristics of Included Studies

Table S4 in Multimedia Appendix 1 summarizes the characteristics of the included studies. All 14 studies were published between 2014 and 2023 and included a total of 5714 participants. About 9 studies (64.3%) were RCTs, with usual care control (n=4, 28.6%), active comparator (n=1, 7.1%), both usual care and active controls (n=2, 14.3%), and waitlist control (n=2, 14.3%). The remaining 5 studies (35.7%) were pre-post studies.

Over half of the studies were conducted in the United States (n=8), followed by the United Kingdom (n=2). The remaining studies were conducted in the United Kingdom and India, New Zealand, Singapore, and across Sweden, South Africa, and Uganda (n=1 each). The sample size across the 14 studies ranged from 27 to 2390 participants. The mean age of the participants was 53.1 years (SD 10.2), with mean ages across studies ranging from 41.7 to 62.4 years (age range: 18 - 88 y). Of the studies that reported the duration of prediabetes (n=4), the duration ranged from 1 to 5 years.

The duration of interventions ranged from 3 to 36 months, with most studies implementing either 6-month or 12-month interventions (n=5 each, 35.7%). Of the 14 included studies, 9 (64.3%) reported follow-up data beyond the endpoint of the intervention, with follow-up periods ranging from 3 to 48 months.

Primary Effectiveness Outcomes

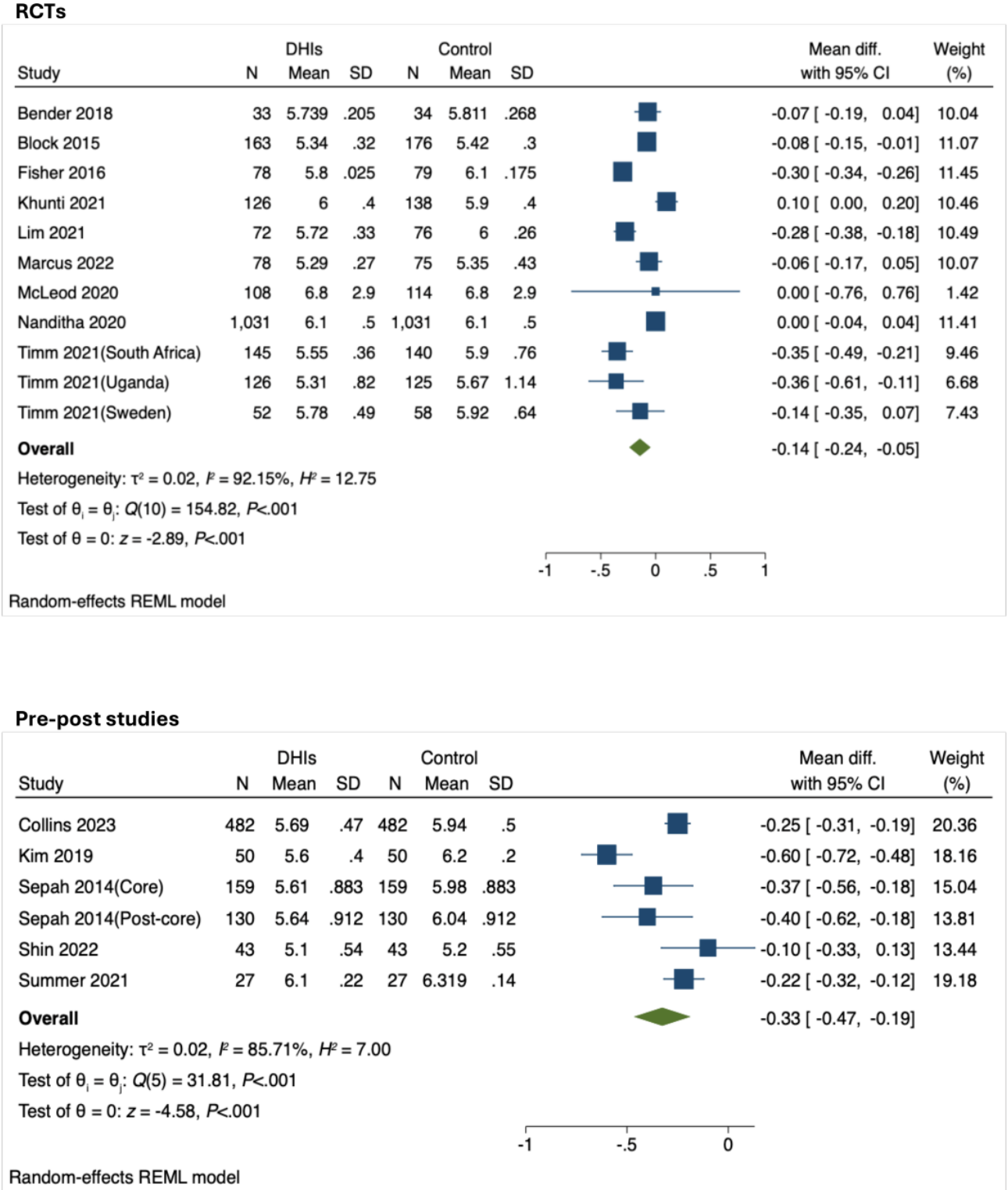
As Cochrane recommends that randomized trials and nonrandomized studies of interventions should not be combined in a meta-analysis [29], the meta-analyses for RCTs and pre-post studies were conducted separately.

Effects of DHIs on HbA_{1c} Levels

The meta-analysis of RCTs (9 studies with 11 datasets, 4058 participants) showed a small effect on HbA_{1c} reduction (mean difference [MD]=−0.14, 95% CI −0.24 to −0.05, P<.001, I^2=92.15%) favoring DHIs (Figure 2). Pre-post studies (5 studies with 6 datasets, 1782 participants) demonstrated a moderate effect on HbA_{1c} levels post-intervention compared to pre-intervention (MD=−0.33, 95% CI −0.47 to −0.19], P<.001, I^2=85.71%; Figure 2). However, the substantial heterogeneity observed limits the interpretability and generalizability of these findings, warranting cautious interpretation.

XSL•FO
RenderX

Figure 2. Forest plots of the effect of digital health interventions (DHIs) on hemoglobin A_{1c} (HbA_{1c}, %) among culturally and linguistically diverse (CaLD) populations living with prediabetes [34-47]. RCTs: randomized controlled trials; REML: restricted maximum likelihood.



Sensitivity Analysis

The robustness of the meta-analysis regarding HbA_{1c} levels was examined through a leave-one-out sensitivity analysis, which was produced by excluding a single study to investigate the influence of each study on the overall effect size estimate. The results in Figure S1 in Multimedia Appendix 2 showed that the pooled effect size remained stable (MD ranging from −0.31 to −0.10), with all *P* values <.001, suggesting the robustness of our findings.

At least 2 trials were identified as a potential source of heterogeneity [35,36]. Removing these studies substantially reduced I^2 in both RCTs (92.15%-89.96%) [32] and pre-post studies (85.71%-0.02% [36]; Figure S2 in Multimedia Appendix 2). The statistical significance of pooled effect sizes remained unchanged (all *P*<.001), indicating that while these studies

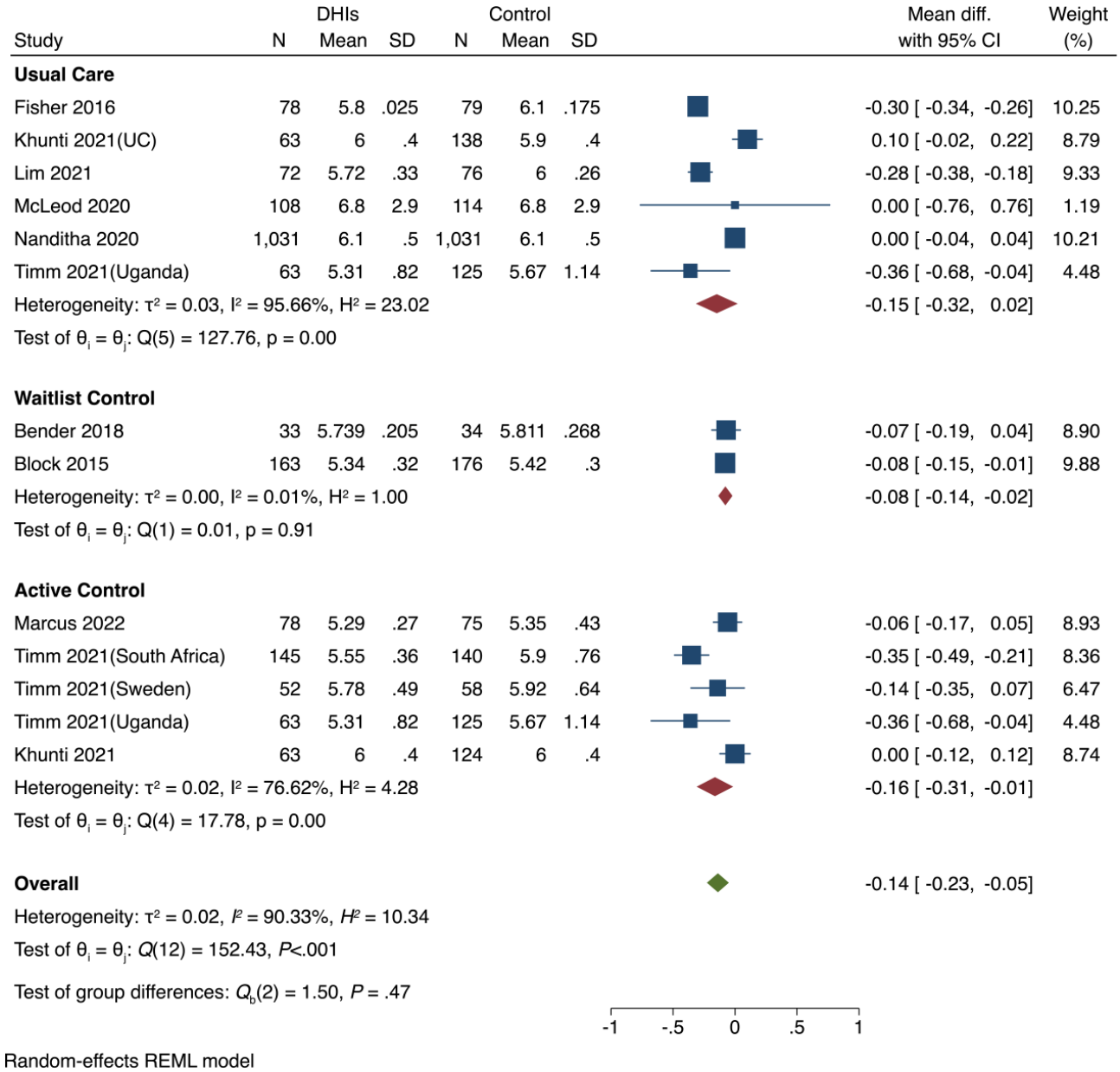
XSL•FO
RenderX

contributed to between-study heterogeneity, they did not alter the significance of the intervention effects.

Subgroup Analysis

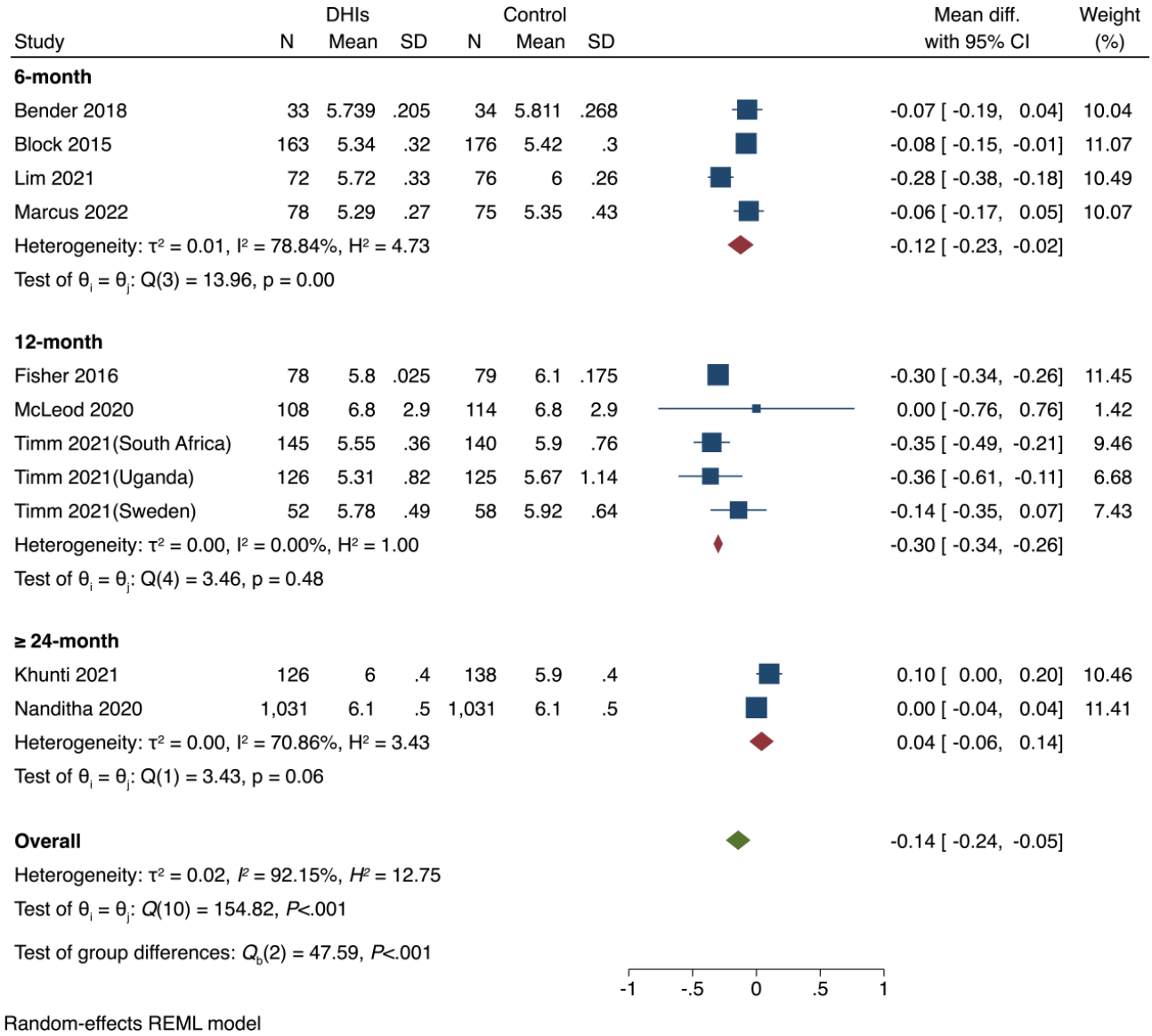
We conducted subgroup analyses of the included RCTs based on control group type and duration of DHIs (6, 12, or 24 mo), which revealed varying intervention effects across different groups (Figures 3 and 4). Figure 3 indicates that both the usual care and active control subgroups showed significant reductions in HbA_{1c} levels (P<.001 for both), while the waitlist control subgroup did not report a significant difference (P=.91). Overall, there were no significant differences among the 3 categories of the control subgroups (P=.47), indicating that the type of the control group did not significantly influence the effect of DHIs.

Figure 3. Subgroup analysis of hemoglobin A_{1c} (HbA_{1c}, %) by control group types [34,35,37-43]. DHIs: digital health interventions; REML: restricted maximum likelihood.



XSL•FO
RenderX

Figure 4. Subgroup analysis of hemoglobin A_{1c} (HbA_{1c}, %) by the duration of digital health interventions (DHIs) [34,35,37-43]. REML: restricted maximum likelihood.



In the subgroup analysis of duration of DHIs in Figure 4, only the 6-month intervention group demonstrated a significant reduction in HbA_{1c} (P<.001). The test for subgroup differences indicates that the effectiveness of DHIs varies significantly across different intervention durations (P<.001). This highlights the importance of careful consideration of intervention duration in the design of DHIs.

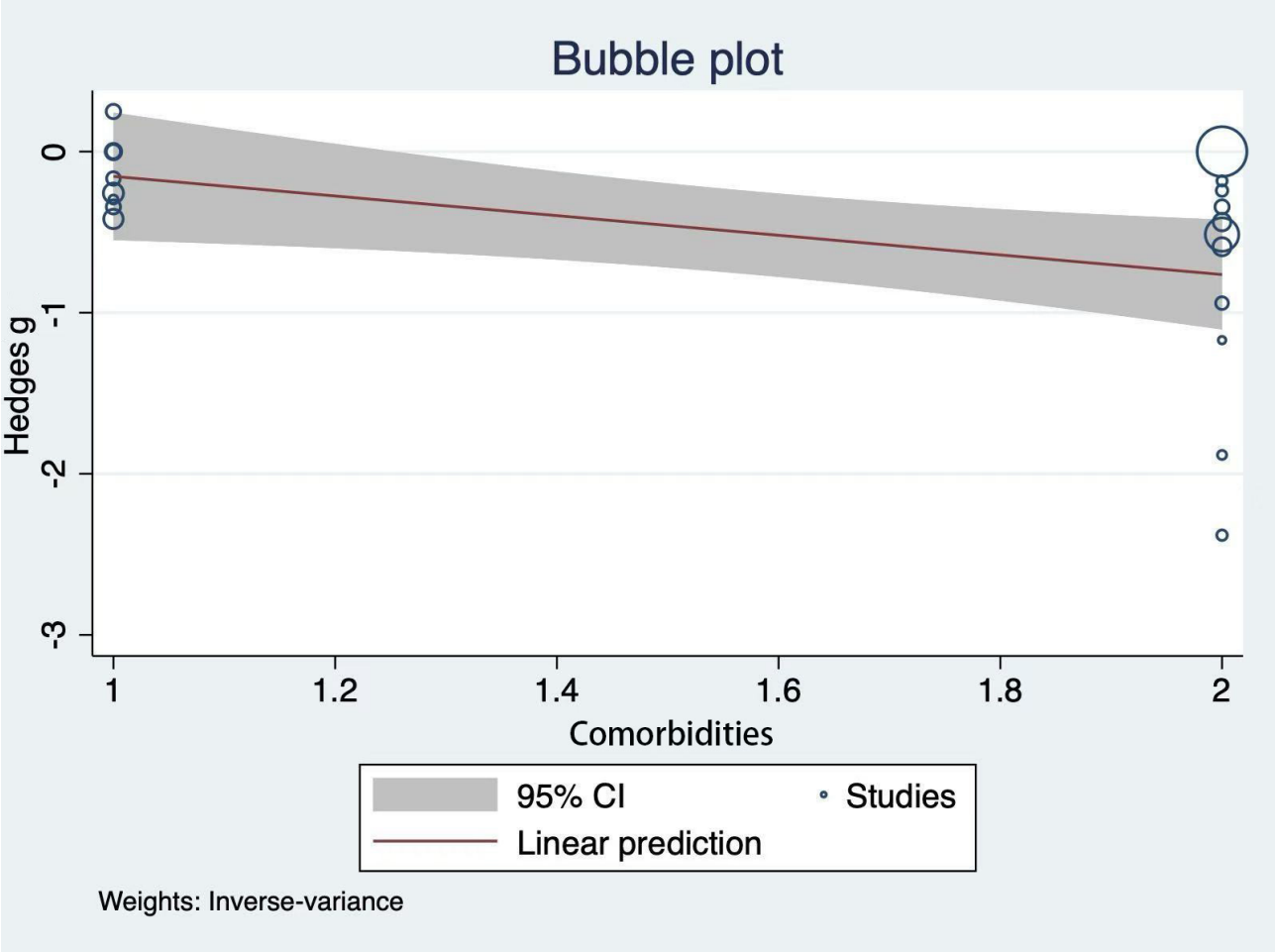
Heterogeneity was observed to be low in most subgroups compared to the overall heterogeneity (Figures 3 and 4). For example, the 12-month intervention group showed no heterogeneity (I^2=0.00%), and the waitlist control group displayed minimal heterogeneity (I^2=0.01%). The decrease in within-group heterogeneity suggests that the duration of DHIs and control group type may be potential sources of heterogeneity when observed across all studies.

Meta-Regression Analysis

A meta-regression analysis was performed to explore the sources of heterogeneity for the outcome HbA_{1c} (I^2=96.23%). Among all covariates tested, including country, sample size, comorbidity status, study design, type of intervention, duration of DHIs, and type of control, only comorbidity status was statistically associated with heterogeneity (P=.02), explaining 20.14% of the observed between-study variance (R^2=20.14%). As shown in the bubble plot in Figure 5, the regression line indicates a negative relationship between comorbidity status and effect size, suggesting that studies reporting participants living with a higher number of comorbidities tend to report smaller intervention effects. However, the wide 95% CI and dispersed study distribution indicate considerable variability in this relationship, limiting the conclusiveness of this association.

XSL•FO
RenderX

Figure 5. Meta-regression bubble plot showing the relationship between participants' comorbidity status and intervention effects on hemoglobin A_{1c} (HbA_{1c}, %).



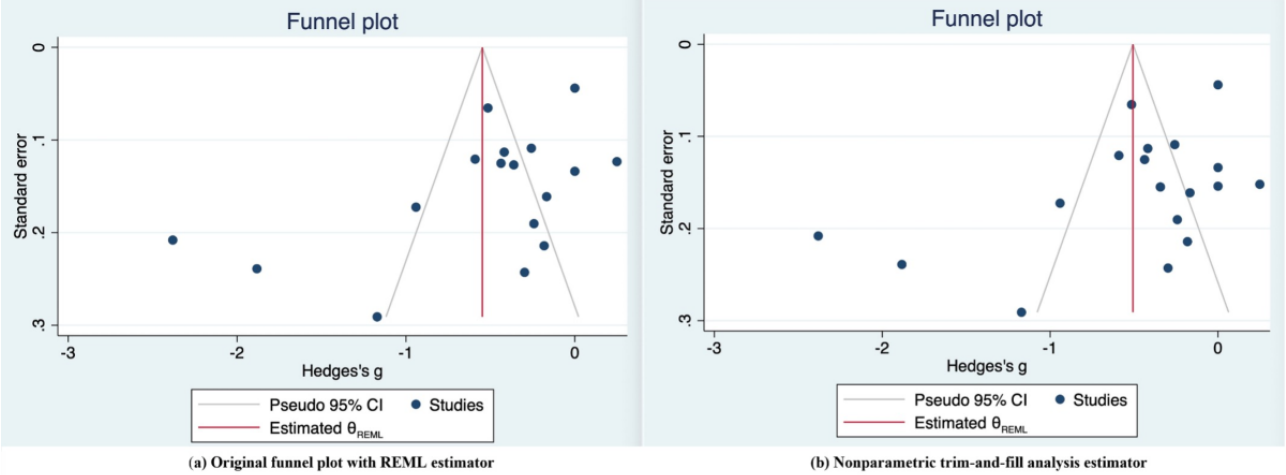
Publication Bias

Publication bias was assessed using both Begg test and Egger test for HbA_{1c} outcomes. While the Begg test suggested no publication bias (*P*=.23), the Egger test indicated potential publication bias (*P*=.02). We also conducted a trim-and-fill analysis to assess the robustness of our findings, in light of the higher sensitivity of the Egger test and the variety of study designs included.

The trim-and-fill analysis identified no missing studies (observed studies=17, imputed studies=0), with the pooled effect size remaining unchanged (95% CI −0.864 to −0.232), suggesting the robustness of the findings. Besides, a visual comparison of the funnel plots in Figure 6 showed that while some asymmetry in study distribution was present, the nonparametric trim-and-fill analysis did not identify any potentially missing studies, as both plots were identical. This suggests that although statistical asymmetry exists, it does not represent true publication bias and therefore does not impact the overall effect estimate.

XSL•FO
RenderX

Figure 6. Publication bias assessment using funnel plots. REML: restricted maximum likelihood.



Secondary Effectiveness Outcomes

Meta-analyses were conducted for FPG and body weight. RCTs showed improvements favoring DHIs for FPG (4 studies, MD=−0.25, 95% CI −0.55 to 0.04, *P*<.001) and body weight (7 studies, MD=−0.30, 95% CI −0.78 to 0.18, *P*<.001). Pre-post studies demonstrated a reduction in body weight (3 studies with 4 datasets, MD=−0.81, 95% CI −2.16 to −0.53, *P*<.001; Figure S3 in Multimedia Appendix 2). Other secondary outcomes, including body mass index, waist circumference, and self-efficacy, had insufficient data for meta-analysis.

Implementation Outcomes

Reach, Uptake, and Engagement

As shown in Table 1, 9 studies (64.3%) reported intervention reach ranging from 11.9% to 86.6%, and 7 studies reported intervention uptake varying from 78.8% to 100%. Among the 14 included studies, participant engagement was reported in 12 studies, with 5 studies reporting digital platform utilization and 3 studies reporting sustained engagement. Participant engagement varied across interventions, with program completion rates ranging from 66.7% (≥75% core lessons) to 100% (≥40% lessons) and session attendance rates from 80% to 96.5%. Detailed implementation metrics are available in Table S5 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Implementation outcomes of digital health interventions.

Author, year	Intervention	Reach (%)	Uptake (%)	Engagement (%)	Retention[a] (%)
Block et al [37] (2015)	Web-based program	NR[b]	NR	70.8	86.1
Bender et al [38] (2018)	Fit and trim program	NR	78.8	91 (visit attendance)	91
Fischer et al [39] (2016)	Message-augmented intervention	14.6	NR	NR	96.3
Lim et al [40] (2021)	App-based coaching	76.4	100	91.7 (app use)	93.2
Marcus et al [41] (2022)	Enhanced PA[c]	21.4	100	50 (sustained)	62.7
McLeod et al [42] (2020)	BetaMe/Melon digital program	17.6	75.9	74 overall (lower for Māori)	94.4
Nanditha et al [35] (2020)	SMS-based lifestyle support	12.3	NR	NR (2 - 3 messages per week)	82
Timm et al [43] (2021)	Telephone-facilitated coaching	2.8 - 35.7	NR	NR	49.8 - 57.6
Khunti et al [34] (2021)	mHealth[d] support	11.9	87.4	80 (group attendance)	73.6
Collins et al [44] (2023)	DPP	NR	NR	16.6 - 98.4 (across behaviors)	46.9
Kim et al [36] (2019)	Mobile self-help program	79.4	100	96.5 (counseling attendance)	87
Shin et al [45] (2022)	Technology-enhanced PA	NR	NR	93 (device adherence)	100
Summers et al [46] (2021)	Low Carb Program	45	100	66.7 (core lesson)	77.8
Sepah et al [47] (2014)	DPP[e]-based group intervention	86.6	100	65.5 (program completion)	65.5

[a]Retention rates reflect the final follow-up point for each study.

[b]NR: not reported.

[c]PA: physical activity.

[d]mHealth: mobile health.

[e]DPP: diabetes prevention program.

Feasibility of DHIs

About 13 studies reported outcomes related to feasibility. Retention rates ranged from 46.9% to 100% across the time points (3 mo: 89.1% - 95.9%, 6 mo: 68.6% - 93.2%, 12 mo: 62.7% - 77.8%), while attrition rates varied from 3.7% to 53.1%. Interestingly, program completion was related to participant characteristics (gender, age, education level), with implementation challenges primarily related to language barriers and delivery methods [34,43,47].

Cultural Adaptation Strategies

Cultural adaptation strategies were reported in 4 studies, including 2 RCTs (with effect sizes of −0.06 and −0.07, respectively) [38,41] and 2 pre-post studies (with effect sizes of −0.10 and −0.60, respectively; Figure 2) [36,45]. These strategies varied across studies and were grouped into 3 main domains. Linguistic adaptations included native language delivery (Korean, Spanish) [36,41,45]. Content-based adaptations involved culturally adapted psycho-behavioral education addressing unique cultural and motivational factors [36], cultural adaptation of theories (social cognitive theory adapted for Latino culture) [41], and integration of culture-specific food guides and activities [38].

Delivery-mode adaptations included personalized tailoring through individual reports and text messages [36,41], as well as community-engaged approaches, such as small group formats, community health worker involvement, and family-centered methods [36,38,45]. The details of these interventions are provided in Table S6 in Multimedia Appendix 1.

GRADE Summary of Findings

Based on the SoF in Table 2, DHIs can result in a slight reduction in HbA_{1c} in RCTs (low-certainty evidence) and may reduce HbA1c in pre-post studies (very low-certainty evidence). For secondary outcomes, DHIs may result in little-to-no difference in FPG (MD=−0.25 mmol, 95% CI −0.55 to 0.04; low-certainty evidence) and body weight (MD=−0.3 kg, 95% CI [−0.78, 0.18]; low-certainty evidence) in RCTs. For pre-post

XSL•FO
RenderX

studies, DHIs may have little-to-no effect on body weight, but the evidence is very uncertain (MD=−0.81 kg, 95% CI −2.16 to 0.53; very low-certainty evidence). The results of FPG and body weight were derived from Figure S3 in Multimedia Appendix 2.

Table . Summary of findings: effectiveness of digital health interventions (DHIs) for culturally and linguistically diverse (CaLD) populations living with prediabetes.

Outcomes	Anticipated absolute effects[a] (95% CI)		Number of participants (studies)	Certainty of the evidence[b] (GRADE[c])	Comments
	Risk with control	Risk with DHIs			
HbA_{1c}[d] follow-up: range 3 months to 48 months	The mean HbA_{1c} ranged from 5.35% to 6.8%	MD[e] 0.14% lower (0.24 lower to 0.05 lower)	4058 (11 RCTs)[f]	Low[g,h]	DHIs may result in a slight reduction in HbA_{1c}
HbA_{1c} follow-up: range 3 months to 12 months	The mean HbA_{1c} ranged from 5.2% to 6.319%	MD 0.33% lower (0.47 lower to 0.19 lower)	891 (6 nonrandomized studies)	Very low[g,i,j,k]	DHIs may reduce HbA_{1c}, but the evidence is very uncertain
FPG[l] follow-up: range 3 months to 24 months	The mean FPG was 5.07 - 6.24 mmol/L	MD 0.25 mmol/L lower (0.55 lower to 0.04 higher)	2616 (4 RCTs)	Low[g,m]	DHIs may result in little-to-no difference in FPG
Body weight follow-up: range 6 months to 48 months	The mean body weight ranged from 78.9 to 92.04 kg	MD 0.3 kg lower (0.78 lower to 0.18 higher)	3575 (7 RCTs)	Low[g,h]	DHIs may result in little-to-no difference in body weight
Body weight follow-up: range 6 months to 12 months	The mean body weight ranged from 66.9 to 99 kg	MD 0.81 kg lower (2.16 lower to 0.53 higher)	1242 (5 nonrandomized studies)	Very low[g,h,i,n]	DHIs may have little-to-no effect on body weight, but the evidence is very uncertain

[a]The risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[b]GRADE Working Group grades of evidence High certainty: we are very confident that the true effect lies close to that of the estimate of the effect. Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect. Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[c]GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]MD: mean difference.

[f]RCTs: randomized controlled trials.

[g]Downgraded 1 level for inconsistency due to substantial statistical heterogeneity (I^2>80%, P<.001), although CIs showed overlap across studies.

[h]The evidence was downgraded 1 level for imprecision due to wide CIs spanning both benefit and harm.

[i]Downgraded 1 level for study limitations due to methodological concerns identified by the JBI critical appraisal tool, particularly regarding insufficient follow-up time (0% compliance) and inadequate reporting of loss to follow-up (20% compliance), although other methodological aspects were well addressed.

[j]Despite being pre-post studies, the evidence directly addressed our research question in terms of population, intervention, and outcomes. Therefore, we did not downgrade for indirectness.

[k]Evidence was downgraded 1 level for imprecision due to 1 study (16.17% weight) having CIs crossing the null line, which impacts the certainty of the effect estimate.

[l]FPG: fasting plasma glucose.

[m]Evidence was downgraded 1 level for imprecision due to CIs crossing the null line in 2 of the 4 studies, despite an adequate total sample size.

[n]Downgraded 2 levels for very serious inconsistency due to substantial statistical heterogeneity (I^2>90%, P<.001) and lack of overlap in CIs across most studies.

Discussion

Principal Findings

This systematic review and meta-analysis examined 14 studies involving 17 articles on the effectiveness, reach, uptake, and feasibility of single- or multi-component DHIs for 5714 participants from CaLD backgrounds living with prediabetes. The meta-analysis indicated that DHIs were effective in improving glycemic control compared to usual care or active controls in both RCTs and pre-post studies (P<.001). However, the evidence from the SoF tables indicates low-to-very low certainty, suggesting that while DHIs may slightly improve HbA_{1c}, the true effect may differ significantly. Similarly, the effects on FPG and body weight are uncertain. In light of the low-to-very low certainty evidence, DHIs are best used as adjunctive tools in prediabetes management with health care provider involvement, rather than as stand-alone solutions, to

XSL•FO
RenderX

maximize effectiveness [6,48]. Regarding intervention implementation, the included studies demonstrated high uptake and strong engagement with DHIs among CaLD populations living with prediabetes, along with satisfactory completion and attendance rates.

Findings in Context

Our meta-analysis found that DHIs, whether stand-alone or integrated, are effective compared to usual care or active control in managing glycemic control in prediabetes. The findings are congruent with previous reviews [20,49], recently published and focused on all populations, although the consistent results are partly due to the overlap of the included studies. They included 3 studies on HbA1c outcomes, 2 of which were also included in our review. We excluded the other study because it was not related to CaLD populations [50]. This situation potentially reflects the limited evidence, particularly from RCTs, on the effectiveness of DHIs in prediabetes management. Our review of 14 studies involving 11 datasets of RCTs provides more comprehensive and systematic evidence, while also extending the evidence specifically to CaLD populations with prediabetes. It should be noted that the inclusion of 5 pre-post studies resulted in very low certainty evidence for both HbA_{1c} and body weight outcomes, primarily due to inherent design limitations such as temporal confounding and lack of control groups.

Several aspects affect the effectiveness of DHIs in managing blood glucose among CaLD populations living with prediabetes. Our subgroup analysis findings suggest that the duration of intervention should be carefully considered when designing DHIs. The results showed that only 6-month interventions had a significant reduction in HbA_{1c} levels (Figure 4). This pattern aligns with previous evidence showing that intervention effects may be difficult to sustain over time. The Finnish Diabetes Prevention Study demonstrated significant improvements in HbA_{1c} at year 1, with benefits diminishing after year 2 [51]. Similarly, mobile health interventions more generally showed that the impact decreased over time, especially after a 6-month follow-up [52]. One possible explanation for this pattern is the decrease in the retention rates and engagement with DHIs across the time points. Participant retention rates reported across 13 studies showed a notable decline, decreasing from high levels at 3 months to moderate levels at 12 months. Engagement data showed similar temporal patterns; for example, Collins et al [44] noted a drop from 74.0% to 46.9%. Additionally, attrition rates showed greater variability in 12-month studies (range: 3.7% - 53.1%) compared to 6-month studies (range: 6.9% - 15.2%; Table S5 in Multimedia Appendix 1).

Moreover, the quality of implementation and cultural adaptation may explain the variations in the effectiveness observed. The outcomes reported at the 12-month follow-up showed lower consistency in effects compared to the 6-month follow-up data, with nonsignificant pooled results due to 2 studies finding null effects (Figure 4) [42,43]. Timm et al. [43] found that within a multicountry intervention, the Sweden subgroup showed no effect due to language barriers and inconsistent delivery of the intervention, while significant effects were noted in the South African and Ugandan subgroups. Additionally, McLeod et al. [42] observed poor engagement among Māori participants, attributing this challenge to the insufficient involvement of Māori and Pacific populations in the design phases. These findings suggest that support for participants to continue to engage with the elements of the intervention is required. In this context, the follow-up at 6 months indicated an optimal balance between sustained engagement and effectiveness. Only 2 studies examined the impact of interventions at 24 months or longer, indicating the need to further understand the longer-term impact of interventions and how participants can be supported to enhance engagement over the longer term.

The meta-regression data indicated a significant association between comorbidity status and the effectiveness of DHIs. This finding aligns with the Look AHEAD trial, which shows that multimorbidity can reduce the effectiveness of lifestyle interventions for diabetes by creating competing treatment demands and increasing the complexity of self-care [53,54]. However, our findings should be interpreted with caution due to the limited explanatory power (R^2=20.14%, P=.02) and require validation in larger studies before informing the design of interventions. Future research should focus on identifying effect moderators and developing individualized, patient-centered DHIs that incorporate disease monitoring and comorbidity management for those living with multiple comorbidities (eg, cardiovascular disease, hyperlipidemia, arthritis) rather than a “one-size-fits-all” approach. The control of confounding factors is essential in intervention studies. Only one of the included studies reported contamination, where both the intervention and control groups received similar additional treatments (eg, diabetes prevention program classes and weight loss programs) [39]. The consideration and reporting of confounding factors need to be addressed when designing future DHIs.

Implementation outcomes for DHIs among CaLD populations living with prediabetes are encouraging, despite high heterogeneity (Table S5 in Multimedia Appendix 1). Qualitative data from 2 RCTs identified key challenges, including language barriers and delivery methods, in implementing DHIs and emphasized the need to tailor interventions [34,43]. These findings are consistent with those from a previous qualitative systematic review, which emphasized the necessity to incorporate cultural and linguistic perspectives into the design and delivery of DHIs to enhance acceptability, appropriateness, and accessibility for underserved populations [9]. Despite this recognized importance, studies employing cultural adaptation strategies remain limited.

In our review, we identified only 4 studies that integrated cultural adaptations into DHIs, specifically involving Korean Americans, Filipino Americans, and Spanish-speaking Latin Americans [36,38,41,45]. Cultural adaptation in this context refers to comprehensive modifications in content and delivery methods, rather than merely translating language. One study, for example, tailored its DHI for Filipino Americans by incorporating cultural considerations into a mobile- and social media–based program. This included, but was not limited to, how to make healthier Filipino meals, providing a Filipino food guide, and promoting indoor and outdoor activities that reflect Filipino culture [38]. Our findings indicated that studies varied

XSL•FO
RenderX

in both the use of cultural adaptation and the specific strategies employed, which may contribute to the observed heterogeneity in HbA_{1c} outcomes. Although subgroup analyses were not feasible with only 2 RCTs employing diverse culturally adapted approaches, the identified strategies provide potential directions for a future intervention design. Evidence supports a tailored approach to the development of information, resources, and interventions for CaLD populations to leverage community resources and expertise and ensure interventions are accessible and relevant to CaLD communities [55]. Researchers and clinicians need to be aware of the importance of creating information, resources, and interventions that are both accessible in terms of language and comprehension and are culturally relevant. Cultural adaptation requires a multilevel approach that simultaneously addresses language, social structures, and practical cultural elements.

Implications for Future Research

Several knowledge gaps need to be addressed in future research. More rigorous research is needed not only to validate the effectiveness of DHIs but also to understand how DHIs can be optimally designed and implemented across diverse populations [37,39,40]. While our review found that most studies (n=13, 92.9%) employed integrated digital approaches, there is a need for future research to consider person-centered and culturally tailored content in addition to focusing on the impact of digital delivery modes. Further research should examine how contextual factors such as gender, age, and comorbidity influence outcomes [34,35]. Additionally, more research is needed to understand how various cultural adaptation strategies impact the effectiveness and implementation of interventions. This will help identify which components are most effective for managing prediabetes.

There is also a critical need to optimize participant engagement strategies, including recruitment methods, duration of DHIs, and follow-up periods, to improve recruitment, completion, and adherence and further enhance the reliability of research findings. Notably, although this review aimed to include people of all ages with prediabetes, all the included studies focused on adults (≥18 y), highlighting a research gap in interventions designed for children and young people living with prediabetes. Future research should prioritize the development and evaluation of culturally adapted DHIs across the entire lifespan, particularly for pediatric and adolescent populations, where prediabetes prevalence has reached alarming levels globally and intensive lifestyle modification is critically needed [56].

Strengths

The strengths of this review included the use of rigorous systematic review and meta-analysis guidelines, separate meta-analyses for RCTs and pre-post studies, the restricted maximum likelihood method for robust random-effects analyses, comprehensive heterogeneity exploration through subgroup analyses and meta-regression, leave-one-out sensitivity analyses to assess robustness, a trim-and-fill method for publication bias to ensure reliability, the inclusion of both effectiveness and implementation metrics, and a SoF table to assess the certainty of evidence.

Limitations

Our review also has several limitations. The high heterogeneity (I^2>75%) and predominantly low-to-very low GRADE certainty substantially limit the interpretability of our findings, which should be considered preliminary and require validation through larger, higher-quality studies. Lack of blinding in the included RCTs may have introduced performance bias. However, the objective nature of our effectiveness outcomes (eg, HbA_{1c}, FPG) and implementation outcomes (eg, retention rates) helps limit measurement bias. Despite our comprehensive, age-inclusive search strategy, no eligible studies were identified targeting children and adolescents from CALD backgrounds, revealing a critical evidence gap in this field. Additionally, most studies were conducted in high-income countries, limiting generalizability to low-resource settings where CaLD populations may face different barriers to health care access and technology adoption. Finally, HbA_{1c} endpoint data from the Swedish arm of a multicenter RCT [43] were not directly reported in the published article. We obtained the original data file (STATA format) from the clinical trial registry and contacted the corresponding author for verification. Means and standard deviations from the original data were calculated for meta-analysis, following Cochrane guidelines on using the available and reliable original data. The results should be interpreted with caution.

Conclusions

This systematic review and meta-analysis suggests that DHIs may be effective in improving glycemic control; however, the low certainty of the evidence indicates that there may be variations in the true effects, and the impacts on FPG and body weight remain uncertain. The observed heterogeneity may be attributed to differences in duration of DHIs, control type, and comorbidity status, with significant effects observed at the 6-month follow-up time point. Trim-and-fill analysis for publication bias indicated no missing studies and unchanged effect estimates. Implementation metrics showed promising results, with high uptake and engagement among CaLD populations with prediabetes. Future research employing rigorous RCT designs, consideration of contextual factors in intervention design and implementation, and the prioritization of person-centered, culturally tailored content for diverse populations across different geographical locations is needed to enhance the evidence base.

Acknowledgments

We are grateful to Lisa Munro for her invaluable assistance in developing our initial search strategies.

XSL•FO
RenderX

Funding

This study did not receive specific funding from any agency in the public, commercial, or not-for-profit sectors.

Data Availability

All data supporting the findings of this systematic review, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart, list of studies excluded at full-text review, study characteristics, critical appraisal results, and detailed analysis (e.g., forest plots of secondary outcomes), are available in the multimedia appendices. Additional materials are available from the corresponding author upon reasonable request.

The authors affirm their commitment to promoting equity, diversity, and inclusion in the conduct and dissemination of research. This work acknowledges the diverse backgrounds and contributions of all authors and collaborators.

Authors' Contributions

Conceptualization: LW
Data curation: ATB, LW, MZ, WHK
Formal analysis: DA, LW, MZ, WHK
Project administration: LW, MZ
Writing – original draft: ATB, DA, LW, MZ, WHK
Writing – review & editing: ATB, DA, LW, MZ, WHK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy; studies excluded at full-text review; quality assessment; characteristics of included studies; reach, uptake, engagement, and feasibility outcomes of digital health interventions across studies; and summary of cultural adaptation strategies used in digital health interventions.
[DOCX File, 171 KB - diabetes_v11i1e70912_app1.docx]

Multimedia Appendix 2
Leave-one-out sensitivity analyses of HbA1c (%) showing the influence of each study on the pooled effect estimates; Forest plots after removing influential studies; and meta-analysis.
[DOCX File, 908 KB - diabetes_v11i1e70912_app2.docx]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 67 KB - diabetes_v11i1e70912_app3.pdf]

References

1. Alvarez S, Coffey R, Mathias PM, Prediabetes AAM. StatPearls: StatPearls Publishing LLC; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK459332/ [accessed 2026-02-09]
2. Rooney MR, Fang M, Ogurtsova K, et al. Global prevalence of prediabetes. Diabetes Care 2023 Jul 1;46(7):1388-1394. [doi: 10.2337/dc22-2376] [Medline: 37196350]
3. Bailey CJ. Prediabetes: never too early to intervene. Lancet Diabetes Endocrinol 2023 Aug;11(8):529-530. [doi: 10.1016/S2213-8587(23)00152-3] [Medline: 37414070]
4. American Diabetes Association Professional Practice Committee. 3. Prevention or delay of diabetes and associated comorbidities: Standards of Care in Diabetes-2024. Diabetes Care 2024 Jan 1;47(Suppl 1):S43-S51. [doi: 10.2337/dc24-S003] [Medline: 38078581]
5. Dinsmore J, Hannigan C, Smith S, et al. A digital health platform for integrated and Proactive Patient-Centered Multimorbidity Self-management and Care (ProACT): protocol for an action research proof-of-concept trial. JMIR Res Protoc 2021 Dec 15;10(12):e22125. [doi: 10.2196/22125] [Medline: 34914613]
6. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
7. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

XSL•FO
RenderX

8. What is digital health? Food and Drug Administration. 2020. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2026-01-09]
9. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023 Feb 28;25:e42719. [doi: 10.2196/42719] [Medline: 36853742]
10. Bol N, Smit ES, Lustria MLA. Tailored health communication: opportunities and challenges in the digital era. Digit Health 2020;6:2055207620958913. [doi: 10.1177/2055207620958913] [Medline: 33029355]
11. Bian RR, Piatt GA, Sen A, et al. The effect of technology-mediated diabetes prevention interventions on weight: a meta-analysis. J Med Internet Res 2017 Mar 27;19(3):e76. [doi: 10.2196/jmir.4709] [Medline: 28347972]
12. Grock S, Ku JH, Kim J, Moin T. A review of technology-assisted interventions for diabetes prevention. Curr Diab Rep 2017 Sep 23;17(11):107. [doi: 10.1007/s11892-017-0948-2] [Medline: 28942537]
13. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207. [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
14. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
15. Villegas V, Shah A, Manson JE, Tobias DK. Prevention of type 2 diabetes through remotely administered lifestyle programs: a systematic review. Contemp Clin Trials 2022 Aug;119(106817):106817. [doi: 10.1016/j.cct.2022.106817] [Medline: 35691488]
16. Avoke D, Elshafeey A, Weinstein R, Kim CH, Martin SS. Digital health in diabetes and cardiovascular disease. Endocr Res 2024;49(3):124-136. [doi: 10.1080/07435800.2024.2341146] [Medline: 38605594]
17. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
18. Kirley K, Sachdev N. Digital health-supported lifestyle change programs to prevent type 2 diabetes. Diabetes Spectr 2018 Nov;31(4):303-309. [doi: 10.2337/ds18-0019] [Medline: 30510384]
19. Stevens S, Gallagher S, Andrews T, Ashall-Payne L, Humphreys L, Leigh S. The effectiveness of digital health technologies for patients with diabetes mellitus: a systematic review. Front Clin Diabetes Healthc 2022;3(936752):936752. [doi: 10.3389/fcdhc.2022.936752] [Medline: 36992773]
20. Stowell M, Dobson R, Garner K, Baig M, Nehren N, Whittaker R. Digital interventions for self-management of prediabetes: a scoping review. PLOS ONE 2024;19(5):e0303074. [doi: 10.1371/journal.pone.0303074] [Medline: 38728296]
21. Australian Government Department of Health. Australian national diabetes strategy 2021-2030. : Australian Government Department of Health; 2021 URL: https://www.health.gov.au/resources/publications/australian-national-diabetes-strategy-2021-2030?language=en [accessed 2026-01-19]
22. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 9;59(1):6-14. [doi: 10.1093/geront/gny138]
23. Pham TTL, Berecki-Gisolf J, Clapperton A, O'Brien KS, Liu S, Gibson K. Definitions of culturally and linguistically diverse (CALD): a literature review of epidemiological research in Australia. Int J Environ Res Public Health 2021 Jan 16;18(2):737. [doi: 10.3390/ijerph18020737] [Medline: 33467144]
24. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7(64):64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
25. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
26. Munn Z, Aromataris E, Tufanaru C, et al. The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI). Int J Evid Based Healthc 2019 Mar;17(1):36-43. [doi: 10.1097/XEB.0000000000000152] [Medline: 30239357]
27. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health, 2nd ed. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240081949 [accessed 2021-01-19]
28. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
29. Reeves BC, Deeks JJ, Higgins JP, et al. Including non‐randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, editors. Cochrane Handbook for Systematic Reviews of Interventions 2019:595-620. [doi: 10.1002/9781119536604]
30. Barker TH, Stone JC, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid Synth 2023 Mar 1;21(3):494-506. [doi: 10.11124/JBIES-22-00430] [Medline: 36727247]
31. Barker TH, Habibi N, Aromataris E, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. JBI Evid Synth 2024 Mar 1;22(3):378-388. [doi: 10.11124/JBIES-23-00268] [Medline: 38287725]

XSL•FO
RenderX

32. Piggott T, Morgan RL, Cuello-Garcia CA, et al. Grading of Recommendations Assessment, Development, and Evaluations (GRADE) notes: extremely serious, GRADE's terminology for rating down by three levels. J Clin Epidemiol 2020 Apr;120:116-120. [doi: 10.1016/j.jclinepi.2019.11.019] [Medline: 31866468]
33. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
34. Khunti K, Griffin S, Brennan A, et al. Behavioural interventions to promote physical activity in a multiethnic population at high risk of diabetes: PROPELS three-arm RCT. Health Technol Assess 2021 Dec;25(77):1-190. [doi: 10.3310/hta25770] [Medline: 34995176]
35. Nanditha A, Thomson H, Susairaj P, et al. A pragmatic and scalable strategy using mobile technology to promote sustained lifestyle changes to prevent type 2 diabetes in India and the UK: a randomised controlled trial. Diabetologia 2020 Mar;63(3):486-496. [doi: 10.1007/s00125-019-05061-y] [Medline: 31919539]
36. Kim MT, Kim KB, Nguyen TH, et al. Motivating people to sustain healthy lifestyles using persuasive technology: a pilot study of Korean Americans with prediabetes and type 2 diabetes. Patient Educ Couns 2019 Apr;102(4):709-717. [doi: 10.1016/j.pec.2018.10.021] [Medline: 30391298]
37. Block G, Azar KM, Romanelli RJ, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240. [doi: 10.2196/jmir.4897] [Medline: 26499966]
38. Bender MS, Cooper BA, Flowers E, Ma R, Arai S. Filipinos Fit and Trim - A feasible and efficacious DPP-based intervention trial. Contemp Clin Trials Commun 2018 Dec;12(76-84):76-84. [doi: 10.1016/j.conctc.2018.09.004]
39. Fischer HH, Fischer IP, Pereira RI, et al. Text message support for weight loss in patients with prediabetes: a randomized clinical trial. Diabetes Care 2016 Aug;39(8):1364-1370. [doi: 10.2337/dc15-2137] [Medline: 26861922]
40. Lim SL, Ong KW, Johal J, et al. A smartphone app-based lifestyle change program for prediabetes (D'LITE Study) in a multiethnic Asian population: a randomized controlled trial. Front Nutr 2021;8:780567. [doi: 10.3389/fnut.2021.780567] [Medline: 35141265]
41. Marcus BH, Dunsiger S, Pekmezi D, Benitez T, Larsen B, Meyer D. Physical activity outcomes from a randomized trial of a theory- and technology-enhanced intervention for Latinas: the Seamos Activas II study. J Behav Med 2022 Feb;45(1):1-13. [doi: 10.1007/s10865-021-00246-6] [Medline: 34379236]
42. McLeod M, Stanley J, Signal V, et al. Impact of a comprehensive digital health programme on HbA1c and weight after 12 months for people with diabetes and prediabetes: a randomised controlled trial. Diabetologia 2020 Dec;63(12):2559-2570. [doi: 10.1007/s00125-020-05261-x] [Medline: 32886192]
43. Timm L, Karlsson I, Sidney Annerstedt K, et al. Intervention fidelity focusing on interaction between participants and facilitators in a telephone-delivered health coaching intervention for the prevention and management of type 2 diabetes. Nutrients 2021 Oct 28;13(11):3862. [doi: 10.3390/nu13113862] [Medline: 34836116]
44. Collins C, McCallum E, Nordby K, et al. Enrollment, engagement, and effectiveness of a large-scale diabetes prevention program using synchronous distance technology. Diabetology 2023;4(4):574-585. [doi: 10.3390/diabetology4040051]
45. Shin C, Oh KM, Lee M, An K, Sim J. A technology-enhanced physical activity intervention: a feasibility study. Clin Nurs Res 2022 Sep;31(7):1219-1224. [doi: 10.1177/10547738221102272] [Medline: 35686393]
46. Summers C, Tobin S, Unwin D. Evaluation of the low carb program digital intervention for the self-management of type 2 diabetes and prediabetes in an NHS England general practice: single-arm prospective study. JMIR Diabetes 2021 Sep 9;6(3):e25751. [doi: 10.2196/25751] [Medline: 34499035]
47. Sepah SC, Jiang L, Peters AL. Translating the diabetes prevention program into an online social network: validation against CDC standards. Diabetes Educ 2014 Jul;40(4):435-443. [doi: 10.1177/0145721714531339] [Medline: 24723130]
48. Duong T, Olsen Q, Menon A, et al. Digital health interventions to prevent type 2 diabetes mellitus: systematic review. J Med Internet Res 2025 Apr 25;27:e67507. [doi: 10.2196/67507] [Medline: 40280555]
49. Nguyen V, Ara P, Simmons D, Osuagwu UL. The role of digital health technology interventions in the prevention of type 2 diabetes mellitus: a systematic review. Clin Med Insights Endocrinol Diabetes 2024;17:11795514241246419. [doi: 10.1177/11795514241246419] [Medline: 38779330]
50. Katula JA, Dressler EV, Kittel CA, et al. Effects of a digital diabetes prevention program: an RCT. Am J Prev Med 2022 Apr;62(4):567-577. [doi: 10.1016/j.amepre.2021.10.023] [Medline: 35151522]
51. Lindström J, Louheranta A, Mannelin M, et al. The Finnish Diabetes Prevention Study (DPS): lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care 2003 Dec;26(12):3230-3236. [doi: 10.2337/diacare.26.12.3230] [Medline: 14633807]
52. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
53. Gregg EW, Holman N, Sophiea M, et al. Multiple long-term conditions as the next transition in the global diabetes epidemic. Commun Med (Lond) 2025 Feb 14;5(1):42. [doi: 10.1038/s43856-025-00742-9] [Medline: 39953177]
54. Espeland MA, Gaussoin SA, Bahnson J, et al. Impact of an 8‐year intensive lifestyle intervention on an index of multimorbidity. J American Geriatrics Society 2020 Oct;68(10):2249-2256. [doi: 10.1111/jgs.16672]

XSL•FO
RenderX

55. Whitehead L, Kirk D, Chejor P, et al. Interventions, programmes and resources that address culturally and linguistically diverse consumer and carers' cancer information needs: a mixed methods systematic review. BMC Cancer 2025 Apr 2;25(1):599. [doi: 10.1186/s12885-025-13931-5] [Medline: 40175999]
56. Han C, Song Q, Ren Y, Chen X, Jiang X, Hu D. Global prevalence of prediabetes in children and adolescents: a systematic review and meta-analysis. J Diabetes 2022 Jul;14(7):434-441. [doi: 10.1111/1753-0407.13291] [Medline: 35790502]

Abbreviations

CaLD: culturally and linguistically diverse
DHIs: digital health interventions
FPG: fasting plasma glucose
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
HbA_{1c}: hemoglobin A_{1c}
IGT: impaired glucose tolerance
JBI: Joanna Briggs Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SoF: Summary of Findings

Edited by S Li; submitted 06.Jan.2025; peer-reviewed by A Girma, Z Dong; revised version received 08.Oct.2025; accepted 01.Nov.2025; published 19.Feb.2026.

Please cite as:
Whitehead L, Zhang M, Kwok WH, Arabiat D, Towell Barnard A
Effectiveness, Reach, Uptake, and Feasibility of Digital Health Interventions for Culturally and Linguistically Diverse Populations Living With Prediabetes Across the Lifespan: Systematic Review and Meta-Analysis
JMIR Diabetes 2026;11:e70912
URL: https://diabetes.jmir.org/2026/1/e70912
doi:10.2196/70912

© Lisa Whitehead, Min Zhang, Wai Hang Kwok, Diana Arabiat, Amanda Towell Barnard. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Continuous Glucose Monitoring–Derived Metrics and Cardiovascular Risk Among People With Diabetes: Systematic Scoping Review

Helene Bei Thomsen[1,2], MSc; Benjamin Lebiecka-Johansen[1], PhD; Ole Nørgaard[3], MSc; Tue Helms Andersen[3], MSc; Signe Toft Andersen[1,4], MD, PhD; Guy Fagherazzi[5], PhD; Adam Hulman[1,2], PhD; Anders Aasted Isaksen[1], MD, PhD

[1]Steno Diabetes Center Aarhus, Palle Juul-Jensens Boulevard 11, Aarhus, Central Denmark Region, Denmark

[2]Department of Public Health, Aarhus University, Aarhus, Central Denmark Region, Denmark

[3]Danish Diabetes Knowledge Center, Department of Education, Copenhagen University Hospital - Steno Diabetes Center Copenhagen, Herlev, Denmark

[4]Medical Department, Gødstrup Hospital, Herning, Central Denmark Region, Denmark

[5]Department of Precision Health, Deep Digital Phenotyping Research Unit, Luxembourg Institute of Health, Strassen, Luxembourg

Corresponding Author:
Helene Bei Thomsen, MSc
Steno Diabetes Center Aarhus, Palle Juul-Jensens Boulevard 11, Aarhus, Central Denmark Region, Denmark

Abstract

Background: Conventional clinical markers guide cardiovascular risk stratification; however, continuous glucose monitoring (CGM) data remain absent from prediction models. A synthesis of the current literature is needed to clarify the prognostic relevance of CGM data for cardiovascular outcomes in people with diabetes.

Objective: This scoping review aimed to identify published studies examining (1) the associations between glycemic control and cardiovascular outcomes and (2) the predictive value of CGM-derived metrics in cardiovascular risk assessment.

Methods: MEDLINE and Embase were searched from inception to March 11, 2025, for peer-reviewed, original research that included CGM-derived metrics and cardiovascular disease (CVD) outcomes. Two reviewers screened the records independently.

Results: A total of 53 studies were identified. These studies focused on type 1 diabetes, type 2 diabetes, both diabetes types, or prediabetes. Clinical outcomes were examined in 16 studies, while subclinical outcomes were assessed in 40 studies. Of the 53 studies, 47 were cross-sectional studies and 6 were longitudinal studies. All studies were association studies, and 3 included secondary analyses of predictive performance. However, none applied machine learning–based methods. A wide range of CGM-derived metrics and CVD outcomes, both clinical and subclinical, were studied in the literature.

Conclusions: Overall, the findings were inconsistent across studies, and this was likely due to methodological weaknesses such as underpowered analyses. Time-in-range was both the most studied metric and associated with cardiovascular risk in the largest single study. Only the mean amplitude of glycemic excursions was consistently associated with CVD in most studies investigating this metric, when using statistical significance as a pragmatic indicator of consistency across heterogeneous studies. The prognostic value of CGM-derived metrics for CVD outcomes is currently underexplored. Longitudinal prediction studies on clinical CVD outcomes, leveraging the potential of routinely collected CGM data, are needed.

(JMIR Diabetes 2026;11:e89374) doi:10.2196/89374

KEYWORDS

diabetes; blood glucose; cardiovascular disease; cardiovascular risk; continuous glucose monitoring; scoping review; prediction; association

Introduction

Cardiovascular disease (CVD) is the main cause of disability and mortality among people with diabetes [1]. Abundant literature exists on the use of simple clinical measurements for risk prediction models to identify individuals at high risk of developing CVD [2,3]. However, these prediction models are rarely used in clinical practice due to methodological flaws and a lack of external validation [4]. However, exceptions do exist, but they are limited to the use of traditional risk factors such as sex, age, smoking status, and routinely collected biomarkers [5,6].

Progress has been made in developing digital tools and wearable technologies, such as continuous glucose monitoring (CGM) devices, to aid decisions in diabetes management [7,8]. CGM has been shown to be an effective tool for achieving glycemic control [9]. The barriers to CGM usage have mostly been overcome [10], and it is expected that the use of CGM devices

XSL•FO
RenderX

will rise as sensors become less obtrusive and more cost-effective [10,11]. This will lead to the accumulation of a large amount of CGM data that may hold predictive potential for CVD prediction, given advances in artificial intelligence and the established links between glycemic control measured by hemoglobin A_{1c} (HbA_{1c}) and CVD risk [12,13]. The predictive aspect has been overlooked in previous efforts to synthesize evidence on the links between CGM data, including CGM-derived metrics, and CVD complications [14,15]. There should be a focus on the distinction between association and prediction, since biomarkers with strong associations can exhibit modest predictive value for risk stratification in precision medicine [16].

Therefore, the objective of this scoping review was to identify studies focusing on CGM-derived metrics as predictors of CVD and assess associations between glycemic control and CVD risk in people with diabetes.

Methods

Scoping Review Framework and Reporting

This scoping review has been conducted according to the Manual for Evidence Synthesis (Chapter 10 - Scoping reviews) from the Joanna Briggs Institute [17] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Checklist 1) [18]. A detailed study protocol for this scoping review has previously been published, along with a description of any deviations from the original protocol [19].

Concepts and Definitions

In this review, the term "CGM-derived metrics" covers all metrics derived from CGM device data. Blood glucose metrics like HbA_{1c} or measurements from anything other than CGM data will not be included (eg, metrics based on self-monitored blood glucose measurements taken with finger-prick or measurements from blood samples such as HbA_{1c}). CVD outcomes were grouped as either clinical or subclinical. The following outcomes were considered clinical CVD: cardiovascular mortality, major adverse cardiovascular events, coronary artery disease, heart failure, stroke, and peripheral artery disease. Synonymous terms (eg, ischemic heart disease) and clinical events (eg, undergoing coronary artery bypass surgery) were also included.

Subclinical outcome measures were grouped into 6 subcategories: arterial stiffness, flow resistance, arterial wall thickness, arterial wall composition, cardiac and pulse-related measures, and arterial lumen. CVD outcomes did not include broader risk factors nonspecific to cardiovascular risk (eg, age and sex).

Eligibility Criteria

The eligibility criteria are reported in Textbox 1, with further details provided in the review protocol [19].

Textbox 1. Eligibility criteria.

Inclusion criteria

- Human clinical studies including participants with prediabetes or any type of diabetes, except for gestational diabetes, regardless of definitions.
- Peer-reviewed published original articles (including brief reports).
- Studies investigating either:
  - The association between continuous glucose monitoring (CGM)-derived metrics of glycemic control and cardiovascular risk markers or cardiovascular diseases (prevalent or incident).
  - CGM-derived metrics of glycemic control as predictors of cardiovascular risk markers or cardiovascular diseases (prevalent or incident).

Exclusion criteria

- Review articles, editorials, case reports, protocols, conference abstracts, and preprints.
- Animal studies not including any human participants.
- Studies not including metrics derived from CGM device data.
- Studies including CGM-derived metrics as outcomes.
- Studies not focusing on cardiovascular disease outcomes according to our definition, as outlined in the Concepts and Definitions section.
- Studies focusing on pregnant women with any form of diabetes, including gestational diabetes.
- Studies where participants were monitored after surgery or during hospitalization (eg, intensive care unit).
- Language not understood by the authors.

Information Sources and Search

The MEDLINE and Embase databases were searched from inception to March 11, 2025, using a search strategy tested against 13 key articles within the field [12,14,20-30] by an information specialist (ON) and reviewed by another (THA; Multimedia Appendix 1) [31].

Selection of Sources of Evidence

Following the search, all identified citations were collated and uploaded into EPPI Reviewer 6, and duplicates were removed

XSL•FO
RenderX

[32]. A meeting was held after about 5% of all titles and abstracts had been screened to create consistency among the reviewers (HBT, BL-J, AH, and AAI). In the screening phase, 2 independent reviewers screened the titles and abstracts to assess eligibility. When all titles and abstracts had been screened, full-text versions of relevant articles were retrieved and assessed in detail against the eligibility criteria by 2 independent reviewers. The reasons for exclusion during full-text screening were recorded and reported. Any disagreements that arose between the reviewers at any stage of the selection process were resolved through discussion. Disagreements unresolved through discussion were settled by the senior researcher (AH).

After the first screening phase, the software tool citationchaser was used for backward and forward citation searching [31]. The tool was applied to all included studies, and the screening process was repeated until no additional studies were found through backward citation and forward citation searching [19].

Data Charting Process and Data Items

Research questions were predefined and published in the scoping review protocol (Textbox 2) [19], and a corresponding data extraction table was developed based on the PRISMA-ScR checklist [18]. Data were extracted by the first author (HBT) and verified by the last author (AAI).

Textbox 2. Research questions.

1. Is there an association between glycemic control and cardiovascular disease (CVD) risk?
2. Can continuous glucose monitoring (CGM)-derived metrics predict CVD risk?
3. What CGM-derived metrics are used in the literature?
4. Which cardiovascular markers and diseases are included as outcomes in the studies?
5. What characterizes study populations (age, sex, ethnicity, or geographic location)?
6. What study designs are used (eg, longitudinal cohort, randomized controlled trial, and cross-sectional)?
7. How was data collected (eg, clinical trial, epidemiological study, and routinely collected data)?
8. What CGM devices were used?
9. What statistical models were used in the studies?
10. Are the data openly available?
11. Is the code openly available?

Synthesis of Results

Study characteristics were aggregated using descriptive statistics, and a narrative summary accompanied the tabulated results.

We considered the most adjusted models to be more clinically relevant and therefore extracted results only from these in the main population of each study, if studies reported numerous estimates due to multiple adjustment levels, subgroup stratifications, and varied combinations of CGM metrics and CVD outcomes. Results are reported for all studies, both clinical and subclinical, but only the specific details of association studies investigating clinical outcomes have been prioritized and presented in the main text, and evidence from the subclinical studies has been presented in Multimedia Appendix 2. A full list of all the CGM metrics and CVD outcomes found in the literature is presented in Multimedia Appendix 3.

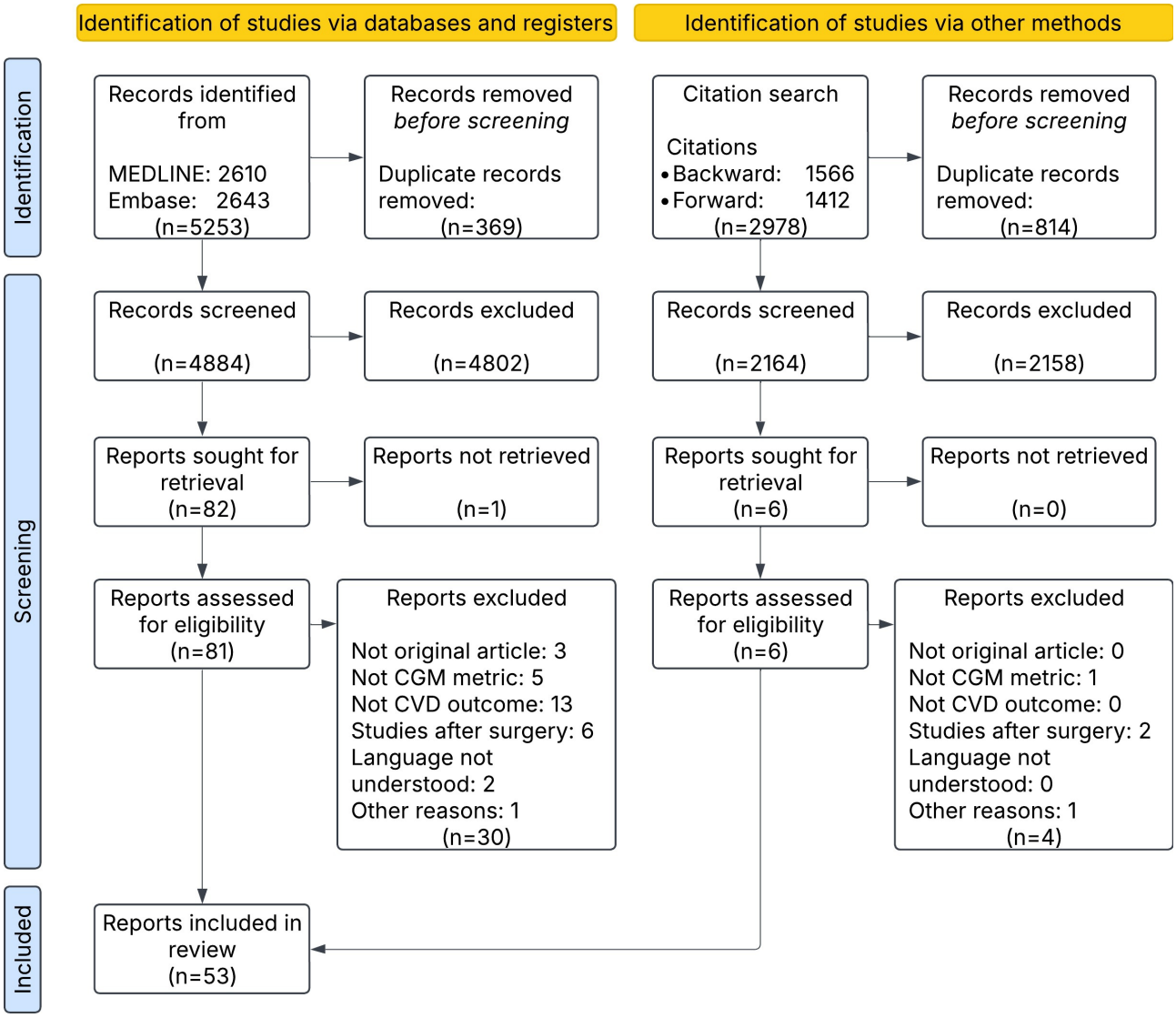
Results

Study Selection Process

The search identified 5253 records, of which 369 were duplicates and therefore removed (Figure 1). After title and abstract screening, 4802 records were excluded, leaving 82 records for full-text screening.

XSL•FO
RenderX

Figure 1. Flow diagram of study selection. CGM: continuous glucose monitoring; CVD: cardiovascular disease.



For full-text screening, 1 study could not be retrieved and 30 studies were excluded for the following reasons: not an original article (n=3), CGM-derived metrics were not based on CGM data (n=5), the patient group did not have CVD as an outcome (n=13), the study was in the postsurgery stage (n=6), the study language was not understood by the authors of this review (n=2), and the study was not yet published in a journal (n=1). Two additional studies were identified through backward and forward citation searching, resulting in a total of 53 included studies on clinical and subclinical outcomes (Multimedia Appendices 4 and 5).

Study Populations

The most common patient group was people with type 2 diabetes (34 out of 53 studies), followed by people with type 1 diabetes (20 out of 53 studies) and those with prediabetes (4 out of 53 studies). The patient group mostly included adults, with 7 studies focusing on children younger than 18 years [33-39]. Almost all studies included both male and female participants, with the exception of 2 studies [40,41], which included only male participants. The geographic distribution of the studies was uneven, as most studies were from Asia (27 out of 53 studies) and Europe (22 out of 53 studies). Three studies included data collected in North America, and 1 study was from Australia. Only 1 study included data from an African country; however, the study population was still predominantly White [42]. No studies were from South America (Multimedia Appendix 6).

Study Designs and Data Collection

One study was a randomized controlled trial [34], and the remaining 52 were observational studies. The majority of studies (49 out of 53 studies) performed cross-sectional analyses, and only 5 studies conducted longitudinal analyses. The size of the study population varied greatly, ranging from 17 to 6225, with a median of 152 (IQR 75 - 469). Eleven studies included routinely collected CGM data from the participants' own devices, and 42 studies used CGM data that had been actively collected with a device provided as part of the study. Studies analyzed data with Spearman or Pearson correlation (n=11), Cox proportional hazards regression (n=3), linear regression (n=23), or logistic regression (n=22; Multimedia Appendices 7 and 8). All studies in this review were association studies, with only 3 studies reporting prediction measures from secondary analyses [28,43,44], and none of these 3 studies used machine learning methods. None of the studies shared data or code.

XSL•FO
RenderX

CGM Findings

Medtronic devices were most frequently used (28 out of 53 studies), followed by Abbott (9 out of 53 studies) and Dexcom (5 out of 53 studies). Other companies were Menarini, Meiqi Company, and SIBIONICS (Multimedia Appendix 7). Sampling frequencies varied from 3 to 15 minutes but were only reported in a minority of studies (21 out of 53 studies).

Among the 53 studies, the most common CGM-derived metrics were time in range (TIR; n=23), mean amplitude of glycemic excursions (MAGE; n=22), mean blood glucose (n=21), SD (n=19), coefficient of variation (CV; n=19), time below range (TBR; n=16), and time above range (TAR; n=15; Multimedia Appendix 9). Most studies did not find a statistically significant association between CGM-derived metrics and CVD. For example, only 7 out of 23 (30%) studies involving TIR found an association between TIR and CVD. However, among studies involving MAGE, a high proportion (14/22, 64%) reported an association between MAGE and CVD. The use of CGM-derived metrics differed among studies involving different diabetes populations. For example, 10 out of 23 (43%) studies involving TIR and 5 out of 22 (23%) studies involving MAGE had populations with type 1 diabetes. Among studies on people with type 1 diabetes, limited studies detected a statistically significant association between CGM-derived metrics and CVD (eg, TIR: 0/10, 0%; MAGE: 2/5, 40%). In contrast, studies on people with type 2 diabetes more often detected a statistically significant association (eg, TIR: 5/11, 45%; MAGE: 10/14, 71%). Studies on people with type 2 diabetes had a larger median sample size (TIR: 510, IQR 405-600; MAGE: 251, IQR 89-411) than studies on people with type 1 diabetes (TIR: 214, IQR 119-547; MAGE: 57, IQR 30-215; Multimedia Appendix 9).

Prediction Studies

Three studies included predictive analyses investigating MAGE as a predictor of CVD using logistic regression models. The reported area under the receiver operating characteristic curve (AUC) was 0.61 in one study [43] and 0.62 in another study [28]. Both studies reported MAGE to be a superior predictor when compared to HbA_{1c}, which had AUC values of 0.55 and 0.58, respectively. In the third study, a receiver operator characteristic curve analysis was conducted to ascertain the optimal threshold for dichotomizing MAGE as part of the variable selection process, but the AUC was not reported [44]. The authors found that MAGE ≥3.4 mmol/L was a risk factor for stenosis and/or occlusion, with a sensitivity of 0.60 and a specificity of 0.61.

Association Studies

Of the 53 included studies, 13 (25%) focused solely on clinical outcomes, 37 (70%) focused solely on subclinical outcomes, and 3 (6%) investigated both outcomes. The design/demographics and main findings of the studies on clinical cardiovascular outcomes are summarized in Tables 1 and 2, respectively, and the results of the studies on subclinical cardiovascular outcomes are summarized in Multimedia Appendix 7.

XSL•FO
RenderX

Table . Study design and demographics of the included studies on clinical cardiovascular outcomes.

Reference	Study design	Population	Size, n	Age[a] (years)	Diabetes duration[b] (years)	HbA_{1c}[c] (mmol/mol, %)	CGM[d] duration
Chen et al [40][e], 2020	Longitudinal (prospective) and cross-sectional (retrospective); FU[f]: in-hospital or within 3 months after discharge from hospital	T2D[g] with CAD[h] (only male); BG[i] control: n=90, BG fluctuation: n=120	210	BG control: mean 55.53 (SD 7.30), BG fluctuation: mean 56.41 (SD 7.67)	BG control: mean 6.59 (SD 2.30), BG fluctuation: mean 6.92 (SD 2.25)	—[j]	2 days
He et al [45], 2023	Longitudinal (prospective); FU: 1 year	T2D with kidney disease on hemodialysis; High TIR[k]: n=12, Low TIR: n=15	27	High TIR: median 66 (IQR 63-73), Low TIR: median 70 (IQR 64-75)	High TIR: median 2 (IQR 1.7-10), Low TIR: median 5 (IQR 0.75-20)	High TIR: median 43 (IQR 37-51) mmol/mol or 6.1% (IQR 5.5%-6.8%), Low TIR: median 66 (IQR 46-70) mmol/mol or 8.2% (IQR 6.4%-8.6%)	14 days
Lu et al [21], 2021	Longitudinal (prospective); FU: until death occurred or 3 - 13 years, median: 6.9 years	T2D; Hospitalized	6225	Mean 61.7	Mean 9.7	Mean 74.0 (SD 24.0) mmol/mol or mean 8.9% (SD 2.2%)	72 hours
Wei et al [46], 2019	Longitudinal (prospective); Median FU: 31 (IQR 22-56) months	T2D; Divided into three groups: (1) No hypoglycemia, n=1173; (2) Mild hypoglycemia (level 1), n=323; (3) Severe hypoglycemia (level 3), n=24	1520	No hypoglycemia: mean 58.59 (SD 11.26), Hypoglycemia: mean 62.27 (SD 11.58)	No hypoglycemia: mean 6.46 (SD 6.00), Hypoglycemia: mean 7.78 (SD 7.37)	No hypoglycemia: mean 8.19% (SD 2.10%), Hypoglycemia: mean 7.73% (SD 1.96%)	3 days
Bezerra et al [47], 2023	Cross-sectional	T1D[l]	161	Mean 37.4 (SD 13.4)	Mean 17.7 (SD 10.6)	Mean 7.5% (SD 1.1%)	14 days
De Meulemeester et al [48][e], 2024	Cross-sectional	T1D	808	Mean 44.8 (SD 15.2)	Mean 23.1 (SD 13.6)	Mean 63 (SD 13) mmol/mol or 7.9% (SD 1.2%)	2 weeks
Deng et al [49], 2023	Cross-sectional	T2D	860	Hp1 carriers: mean 53.5 (SD 13.3), Hp2 - 2: mean 51.7 (SD 14.6)	Hp1 carriers: mean 8.6 (SD 6.4), Hp2 - 2: mean 8.6 (SD 6.7)	Hp1 carriers: mean 73.0 (SD 24.0) mmol/mol or 8.8% (SD 2.2%), Hp2-2: mean 72.0 (SD 23.0) mmol/mol or 8.7% (SD 2.1%)	3 days
El Malahi et al [50], 2022	Cross-sectional	T1D starting on sensor-augmented pump therapy	515	Mean 42.2 (SD 12.5)	Mean 22.3 (SD 11.6)	Mean 60 (SD 9.8) mmol/mol or 7.6% (SD 0.9%)	2 weeks

XSL•FO
RenderX

Reference	Study design	Population	Size, n	Age[a] (years)	Diabetes duration[b] (years)	HbA_{1c}[c] (mmol/mol, %)	CGM[d] duration
Guo et al [51], 2021	Cross-sectional	T1D or T2D with atrial fibrillation; With stroke: n=48, Without stroke: n=462	510	Stroke: mean 70.3 (SD 12.1), No stroke: mean 68.1 (SD 9.4)	—	Stroke: mean 8.2 (SD 1.7), No stroke: mean 7.4 (SD 2.1)	72 hours
Li et al [52], 2020	Cross-sectional	T2D with LEAD[m]: n=179, T2D without LEAD: n=157	336	With LEAD: mean 65.56 (SD 11.99), Without LEAD: mean 55.94 (SD 12.45)	With LEAD: mean 10.32 (SD 4.14), Without LEAD: mean 6.92 (SD 3.54)	With LEAD: mean 8.97% (SD 1.63%), Without LEAD: mean 7.85% (SD 1.41%)	72 hours
Magri et al [27][e], 2018	Cross-sectional	T2D	121	Median 64 (IQR 57-68)	Median 3 (IQR 2-5)	Median 45 mmol/mol (6.8%)	72 hours
Shu-Hua et al [43], 2012	Cross-sectional	T2D with chest pain; Without CAD: n=202, With CAD, n=84	286	Without CAD: mean 62.8 (SD 8.7), With CAD: mean 66.6 (SD 9.2)	—	Without CAD: mean 7.51% (SD 0.80%), With CAD: mean 7.75% (SD 0.92%)	72 hours; Only used the intermediate 48 hours
Sheng et al [53], 2023	Cross-sectional	T2D; Hospitalized	545	Mean 61.22 (SD 11.21)	—	Mean 8.51% (SD 1.85%)	7 - 14 days
Su et al [28], 2011	Cross-sectional	T2D with chest pain; Without CAD: n=92, With CAD: n=252	344	Without CAD: mean 61 (SD 9), With CAD: mean 65 (SD 9)	Without CAD: mean 4.8 (SD 5.7), With CAD: mean 6.5 (SD 6.4)	Without CAD: mean 7.5% (SD 1.4%), With CAD: mean 7.6% (SD 1.5%)	72 hours; Only 48 hours used
Watanabe et al [54], 2017	Cross-sectional	Prediabetes; Hospitalized	28	Mean 64.3 (SD 12.8)	—	Mean 5.41% (SD 0.35%)	72 hours; Only used the middle 48 hours
Zhang et al [30], 2013	Cross-sectional	T2D with cardiovascular complications; Group A: healthy individuals, Group B: T2D without cardiovascular complications, Group C: T2D with cardiovascular complications	92	Group A: mean 56.3 (SD 6.1), Group B: mean 56.1 (SD 6.6), Group C: mean 61.7 (SD 7.2)	—	Group A: mean 5.3% (SD 0.3%), Group B: mean 6.6% (SD 1.2%), Group C: mean 7.5% (SD 1.4%)	72 hours

[a]Age is reported as an interval, mean (SD), or median (IQR).

[b]Diabetes duration values originally reported in months were converted to years for consistency (months ÷ 12).

[c]HbA_{1c}: hemoglobin A_{1c}.

[d]CGM: continuous glucose monitoring.

[e]This study appears in both the clinical and subclinical disease outcome tables owing to the investigation of multiple cardiovascular disease outcomes.

[f]FU: follow-up.

[g]T2D: type 2 diabetes.

[h]CAD: coronary artery disease.

[i]BG: blood glucose.

[j]Not available or not reported.

[k]TIR: time in range.

[l]T1D: type 1 diabetes.

[m]LEAD: lower extremity arterial disease.

XSL•FO
RenderX

Table . Main findings of the included studies on clinical cardiovascular outcomes[a].

Outcome, reference, and continuous glucose monitoring metrics	Unadjusted or least adjusted findings[b]	*P* value for the least adjusted findings	Most adjusted findings[b]	*P* value for the most adjusted findings
Cardiovascular mortality				
Lu et al [21], 2021				
TIR[c] >85%	HR 1.00	<.001 (trend)	HR 1.00	.02 (trend)
TIR 71% - 85%	HR 1.43 (0.95 - 2.14)	<.001 (trend)	HR 1.35 (0.90-2.04)	.02 (trend)
TIR 51% - 70%	HR 1.66 (1.12 - 2.45)	<.001 (trend)	HR 1.47 (0.99-2.19)	.02 (trend)
TIR ≥50%	HR 2.15 (1.47 - 3.13)	<.001 (trend)	HR 1.85 (1.25-2.72)	.02 (trend)
TIR as a continuous variable (each 10% decrease)	HR 1.08 (1.03 - 1.13)	—[d]	HR 1.05 (1.00 - 1.11)	—
Wei et al [46][e], 2019				
Hypoglycemic events	HR 2.033 (1.211-3.413)	—	HR 2.642 (1.398-4.994)	—
Major adverse cardiovascular events				
He et al [45], 2023				
Blood glucose risk index	HR 0.97 (0.85-1.10)	.61	HR 0.98 (0.85-1.13)	.75
Low blood glucose index	HR 2.37 (1.16-4.83)	.02	HR 2.73 (1.21-6.16)	.02
High blood glucose index	HR 0.94 (0.81-1.08)	.38	HR 0.94 (0.81-1.09)	.44
Average of daily risk range	HR 1.00 (0.93-1.07)	>.99	HR 1.01 (0.93-1.09)	.80
GMI[f]	HR 0.98 (0.91-1.06)	.65	HR 0.99 (0.91-1.07)	.78
M-value	HR 0.98 (0.91-1.05)	.54	HR 0.98 (0.91-1.06)	.64
Wei et al [46][e], 2019				
Hypoglycemic events	HR 1.501 (1.207-1.866)	—	HR 1.615 (1.239-2.106)	<.001
Macrovascular complications				
De Meulemeester et al [48][e,g], 2024				
TIR	OR 0.939 (0.829-1.063)	>.05	OR 0.896 (0.738-1.087)	>.05
TITR[h]	OR 0.901 (0.775-1.047)	>.05	OR 0.933 (0.745-1.169)	>.05
Bezerra et al [47], 2023				
TIR	OR 0.66 (0.46 - 0.93)	.02	OR 0.68 (0.39-1.16)	.15
Time below 54 mg/dL	OR 1.10 (0.88 - 1.38)	.39	OR 0.92 (0.62-1.34)	.65
TBR[i]	OR 0.93 (0.80 - 1.09)	.39	OR 0.77 (0.54-1.11)	.17
TAR[j]	OR 1.04 (1.01 - 1.08)	.01	OR 1.04 (0.99-1.10)	.08
Time above 250 mg/dL	OR 1.04 (1.00 - 1.08)	.03	OR 1.03 (0.97-1.09)	.29
CV[k]	OR 1.03 (0.95 - 1.11)	.52	OR 0.92 (0.81-1.06)	.25
GMI	OR 2.17 (1.14 - 4.11)	.02	OR 2.03 (0.77-5.37)	.15
Deng et al [49], 2023				
%CV tertile 1 (Hp1; reference)	OR 1.000	.07 (interaction)	OR 1.000	.008 (interaction)
%CV tertile 1 (Hp2 - 2; reference)	OR 1.000	.07 (interaction)	OR 1.000	.008 (interaction)

XSL•FO
RenderX

Outcome, reference, and continuous glucose monitoring metrics	Unadjusted or least adjusted findings[b]	*P* value for the least adjusted findings	Most adjusted findings[b]	*P* value for the most adjusted findings
%CV tertile 2 (Hp1)	OR 1.483 (0.907-2.423)	.12	OR 1.048 (0.528-2.078)	.89
%CV tertile 2 (Hp2 - 2)	OR 1.399 (0.829-2.358)	.21	OR 0.659 (0.296-1.466)	.31
%CV tertile 3 (Hp1)	OR 2.347 (1.393-3.957)	.001	OR 2.461 (1.183-5.121)	.02
%CV tertile 3 (Hp2 - 2)	OR 1.217 (0.731-2.027)	.45	OR 0.540 (0.245-1.191)	.13
El Malahi et al [50], 2022				
TIR	—	—	—	>.05
SD	—	—	—	>.05
CV	—	—	—	>.05
Magri et al [27][g], 2018				
TBR	—	—	OR 1.12 (1.014 - 1.228)	.02
Lowest BG[l] value	—	—	—	—
Area under the TBR curve	—	—	—	—
Coronary artery disease				
De Meulemeester et al [48][e,g], 2024				
TITR	OR 1.039 (0.812-1.330)	>.05	OR 1.255 (0.874-1.803)	>.05
TIR	OR 1.072 (0.866-1.328)	>.05	OR 1.164 (0.844-1.607)	>.05
Sheng et al [53], 2023				
TIR <20%	—	—	OR 2.143 (1.554 - 3.287)	—
TIR 20 - 40%	—	—	OR 1.049 (0.945 - 2.022)	—
TIR 40 - 60%	—	—	OR 0.854 (0.495 - 1.473)	—
TIR 60 - 80%	—	—	OR 0.617 (0.423 - 1.312)	—
TIR >80%	—	—	OR 0.470 (0.143 - 1.545)	—
Chen et al [40][e,g], 2020				
Controls with SD <1.40 mmol/L, MAGE[m] <3.90 mmol/L, LAGE[n] <4.40 mmol/L, MODD[o] <0.83 mmol/L versus high BG fluctuations (myocardial Infarction)	—	—	^{2}=5.797	.02
Controls with SD <1.40 mmol/L, MAGE <3.90 mmol/L, LAGE <4.40 mmol/L, MODD <0.83 mmol/L versus high BG fluctuations (angina pectoris)	—	—	^{2}=7.490	.006
Wei et al [46][e], 2019				
Hypoglycemic events (myocardial Infarction)	HR 1.901 (1.067-3.389)	—	HR 1.549 (0.768-3.124)	.03
Hypoglycemic events (unstable angina pectoris)	HR 1.226 (0.857-1.753)	—	HR 1.218 (0.794-1.869)	.30
Shu-Hua et al [43][e], 2012				

XSL•FO
RenderX

Outcome, reference, and continuous glucose monitoring metrics	Unadjusted or least adjusted findings[b]	*P* value for the least adjusted findings	Most adjusted findings[b]	*P* value for the most adjusted findings
MAGE level (≥3.4 mmol/L)	—	—	OR 2.286 (1.176-4.446)	.02
Su et al [28][e], 2011				
MAGE ≥3.4 mmol/L	—	—	OR 2.612 (1.423-4.831)	.002
MAGE	—	—	AUC 0.618 (0.555-0.680)	.001
Gensini score				
Chen et al [40][e,g], 2020				
Controls with SD <1.40 mmol/L, MAGE <3.90 mmol/L, LAGE <4.40 mmol/L, MODD <0.83 mmol/L versus high BG fluctuations	—	—	=6.210	<.001
Watanabe et al [54][e], 2017				
MAGE	—	—	*r*=0.742	<.001
Shu-Hua et al [43][e], 2012				
MAGE	—	—	Unstandardized coefficient β=4.817; SE=1.614; standardized coefficient β=0.170; *t*=2.984	.003
Su et al [28][e], 2011				
MAGE	—	—	Unstandardized β=7.010; SE=1.466; standardized β=0.237; *t*=4.783	<.001
Syntax score				
Watanabe et al [54][e], 2017				
MAGE	—	—	*r*=0.776	<.001
Zhang et al [30][e], 2013				
MAGE	—	—	*r*=0.518	.01
BG fluctuations from 00:00 to 03:00	—	—	*r*=−0.442	.04
BG fluctuations from 03:00 to 06:00	—	—	*r*=−0.208	.34
BG fluctuations from 06:00 to 08:00	—	—	*r*=0.678	<.001
BG fluctuations from 08:00 to 11:00	—	—	*r*=0.115	.60
BG fluctuations from 11:00 to 13:00	—	—	*r*=0.523	.01
BG fluctuations from 13:00 to 17:00	—	—	*r*=0.257	.24
BG fluctuations from 17:00 to 19:00	—	—	*r*=0.358	.09
BG fluctuations from 19:00 to 24:00	—	—	*r*=−0.018	.93
Stroke				
De Meulemeester et al [48][e,g], 2024				

XSL•FO
RenderX

Outcome, reference, and continuous glucose monitoring metrics	Unadjusted or least adjusted findings[b]	*P* value for the least adjusted findings	Most adjusted findings[b]	*P* value for the most adjusted findings
TITR	OR 0.651 (0.470-0.902)	<.05	OR 0.546 (0.347-0.858)	<.01
TIR	OR 0.749 (0.588-0.955)	<.05	OR 0.617 (0.440-0.866)	<.01
Guo et al [51], 2021				
TIR: Q1 (≤46%; reference)	OR 1.00	<.001	OR 1.00	<.001
TIR: Q2 (46% - 65%)	OR 0.86 (0.72-0.95)	<.001	OR 0.80 (0.68-0.92)	<.001
TIR: Q3 (65% - 81%)	OR 0.71 (0.61-0.81)	<.001	OR 0.64 (0.53-0.79)	<.001
TIR: Q4 (>81%)	OR 0.66 (0.58-0.80)	<.001	OR 0.59 (0.50-0.74)	<.001
TIR (per 10% increase)	OR 0.93 (0.85-0.98)	.008	OR 0.89 (0.82-0.95)	.001
Wei et al [46][e], 2019				
Hypoglycemic events	HR 1.691 (1.144-2.499)	—	HR 1.813 (1.110-2.960)	.06
Peripheral artery disease				
De Meulemeester et al [48][e,g], 2024				
TITR	OR 0.680 (0.426-1.085)	>.05	OR 0.807 (0.382-1.703)	>.05
TIR	OR 0.736 (0.520-1.042)	>.05	OR 0.811 (0.470-1.398)	>.05
Lower extremity arterial disease				
Li et al [52], 2020				
TIR	OR 0.979 (0.968-0.991)	<.001	OR 0.979 (0.965-0.992)	.002
CV	OR 1.040 (1.003-1.078)	.04	OR 1.038 (0.996-1.081)	.08
SD	OR 1.325 (1.038-1.691)	.02	OR 1.158 (0.824-1.627)	.40
TIR-without LEAD[p] (1)	OR 1.00	—	OR 1.00	—
TIR-mild LEAD (1)	OR 0.98 (0.97-1.00)	.14	OR 0.99 (0.97-1.01)	.25
TIR-moderate LEAD (1)	OR 0.97 (0.95-0.99)	.007	OR 0.97 (0.95-0.99)	.01
TIR-without severe LEAD (1)	OR 0.96 (0.94-0.98)	.002	OR 0.96 (0.94-0.98)	.003
CV-without LEAD	—	—	OR 1.00	—
CV-mild LEAD	—	—	OR 1.03 (0.98-1.07)	.28
CV-moderate LEAD	—	—	OR 1.02 (0.96-1.09)	.48
CV-without severe LEAD	—	—	OR 1.02 (0.95-1.09)	.60
TIR-without LEAD (2)	—	—	OR 1.00	—
TIR-mild LEAD (2)	—	—	OR 0.97 (0.96-1.08)	.06
TIR-moderate LEAD (2)	—	—	OR 0.98 (0.95-0.99)	.01
TIR-without severe LEAD (2)	—	—	OR 0.97 (0.95-0.99)	.02
SD-without LEAD	—	—	OR 1.00	—
SD-mild LEAD	—	—	OR 0.88 (0.47-1.64)	.69
SD-moderate LEAD	—	—	OR 1.28 (0.58-3.07)	.58

XSL•FO
RenderX

Outcome, reference, and continuous glucose monitoring metrics	Unadjusted or least adjusted findings[b]	*P* value for the least adjusted findings	Most adjusted findings[b]	*P* value for the most adjusted findings
SD-without severe LEAD	—	—	OR 1.52 (0.92-2.41)	.10

[a]The full table with adjustments is provided in Multimedia Appendix 10. Further elaboration on the adjusted variables can be found in Multimedia Appendices 7 and 8.

[b]All hazard ratios (HRs), odds ratios (ORs), and areas under the curve (AUCs) are reported as follows: point estimate (95% CI).

[c]TIR: time in range.

[d]Not applicable or not available/not reported.

[e]This study appears multiple times as it investigated multiple cardiovascular disease outcomes.

[f]GMI: glucose management indicator.

[g]This study appears in both the clinical and subclinical disease outcome tables owing to the investigation of multiple cardiovascular disease outcomes.

[h]TITR: time in tight range.

[i]TBR: time below range.

[j]TAR: time above range.

[k]CV: coefficient of variation.

[l]BG: blood glucose.

[m]MAGE: mean amplitude of glycemic excursions.

[n]LAGE: largest amplitude of glycemic excursions.

[o]MODD: mean of daily differences.

[p]LEAD: lower extremity arterial disease.

Cardiovascular Mortality

Two longitudinal studies found an association between a CGM-derived metric (TIR and hypoglycemia) and cardiovascular mortality [21,46].

Major Adverse Cardiovascular Events

Two studies assessed major adverse cardiovascular events [45,46]. Both studies included nonfatal myocardial infarction, nonfatal stroke, and cardiovascular death. One study also included unstable angina leading to hospitalization [46]. Hypoglycemic events and low blood glucose index values were associated with major adverse cardiovascular events; however, no associations were found for other CGM-derived metrics, including glucose management indicator, high blood glucose index, average of daily risk range, m-value, and blood glucose risk index.

Macrovascular Complications

Five studies explored the association between CGM-derived metrics and nonfatal cardiovascular events regardless of anatomical location as a composite CVD outcome, with some variation between the studies in terms of the complications included [27,47-50]. All 5 studies included cerebrovascular accident, 4 included peripheral artery disease [47-50], 3 included coronary artery disease [48-50], 2 included ischemic heart disease [27,47], and 1 each included stenosis [48], heart failure [50], and ankle-brachial index <0.9 or abnormal carotid intima-media thickness [27]. One study [27] found an association between TBR and cardiovascular complications, while another study [47] did not find an association. A study by Deng et al [49] found an association between CV and diabetic macroangiopathy in people who were Hp1 carriers but not in people with the Hp2‑2 genotype. Furthermore, no studies found evidence for associations between macrovascular complications and the following CGM-derived metrics: CV, SD, TIR, TAR, glucose management indicator, time in tight range (TITR), lowest blood glucose value, and area under the TBR curve [27,47,48,50].

Coronary Artery Disease

Six studies investigated coronary artery disease [28,40,43,46,48,53]. Studies reported an association between MAGE ≥3.4 mmol/L and coronary artery disease [28,43]. Furthermore, a difference was observed between the control group and the high blood glucose fluctuation group based on dichotomizing the following CGM metrics: SD <1.40 mmol/L, MAGE <3.90 mmol/L, largest amplitude of glycemic excursions <4.40 mmol/L, and mean of daily differences <0.83 mmol/L [28,43,53]. TIR <20% was found to have an association, but none of the other TIR intervals [53] or TITR [48] had an association. Lastly, hypoglycemic events were found to have an association with myocardial infarction but not with unstable angina pectoris [46]. Four studies investigated the severity of coronary artery disease [55] by using the Gensini score [28,40,43,54]. Three studies found an association with MAGE [28,43,54]. Further associations were found when comparing the control group with the high blood glucose fluctuation group [28,40]. Two studies investigated the complexity of coronary artery disease [56] using the SYNTAX score [30,54]. MAGE was found to have an association with the SYNTAX score [30,54] together with blood glucose excursions during the night from 00:00 to 03:00, in the mornings from 06:00 to 08:00, and at midday from 11:00 to 13:00. No association was detected for all the other times during the day [30].

Stroke

Three studies investigated stroke or cerebrovascular accidents. TIR, but not hypoglycemic events, was associated with stroke

XSL•FO
RenderX

[46,51]. TIR and TITR were both associated with cerebrovascular accidents [48].

Peripheral Artery Disease

A study by De Meulemeester et al [48] included peripheral artery disease, while a study by Li et al [52] included lower extremity artery disease. Overall, they investigated CV, SD, TIR, and TITR, and only an association with TIR was found in some analyses [52].

Discussion

Main Findings and Methodological Considerations

This scoping review identified 53 studies focusing on the relationship between CGM-derived metrics and CVD risk in individuals with diabetes. The literature included inconsistent findings across association studies, which also had highly diverse clinical and subclinical CVD outcomes. CGM-derived metrics are widely studied, but their predictive value for CVD outcomes remains unclear since MAGE was the only metric whose predictive value was tested.

We observed patterns regarding study population size, diabetes type, and reporting of evidence for associations between CGM-derived metrics and CVD outcomes, as studies focusing on type 1 diabetes were often conducted in smaller study populations and rarely found evidence for associations. In contrast, studies focusing on type 2 diabetes were conducted in larger study populations and more consistently found evidence for associations. These patterns may explain some of the inconsistent findings for each CGM-derived metric and could indicate a lack of statistical power in some studies, suggesting that future studies, particularly those focusing on type 1 diabetes, should emphasize having sufficiently sized study populations. However, differences between type 1 and type 2 diabetes populations extend beyond sample size and include distinct pathophysiology, treatment regimens, and cumulative exposure to cardiovascular risk factors, all of which may influence the relationship between CGM-derived metrics and cardiovascular outcomes.

The most frequently investigated CGM-derived metric in our review was TIR, and we found inconsistent results across the included studies. Similar results were reported in the review by Yapanis et al [14]. The authors argued that low TIR is a risk factor for macrovascular disease and mentioned that the large sample size of a supporting study [21] provides more reliable evidence than the inconsistent results reported from smaller sample sizes. The same study [21] was the largest in our review, and its size makes its conclusion compelling. The inconsistency across other TIR studies is likely due to limited power and cross-sectional designs. MAGE was the CGM-derived metric most consistently associated with cardiovascular outcomes across studies and the only CGM-derived metric used for prediction; however, it exhibited poor discriminative ability. The more consistent associations observed for MAGE across studies, despite generally smaller sample sizes compared to TIR, may suggest a stronger link between MAGE and CVD risk than between TIR and CVD risk. Even though MAGE was studied across a diverse range of subclinical outcomes (Multimedia Appendix 7), studies on clinical outcomes were limited to coronary artery disease and severity scores, and studies on other clinical outcomes are needed to confirm this pattern, which would suggest that within-day glycemic variability may be an important cardiovascular risk factor. However, since MAGE is biased toward detecting hyperglycemic excursions [57], it may underestimate the impact of hypoglycemia on CVD. Although some studies have reported associations between hyperglycemia-focused CGM-derived metrics and cardiovascular outcomes [26,43], other studies have reported associations between hypoglycemia and cardiovascular outcomes [27,39,46,58-60], with some concluding that hypoglycemia is associated with macrovascular complications and hyperglycemia is associated with microvascular complications [27,59].

The heterogeneity in the reported associations between CGM-derived metrics and CVD outcomes may also reflect differences in how the metrics were defined and analyzed. Thresholds for TIR, TAR, and TBR, as well as observation periods, varied across studies. Statistical adjustment strategies also differed. Some studies adjusted for HbA_{1c} or other covariates, while others made no adjustments. However, we found no substantial differences between the unadjusted and the most adjusted estimates in the studies, suggesting that attenuation due to adjustment for other covariates played only a minor role. Study comparisons would have been easier if the analysis code were available; however, none of the authors provided this information. Together, these discrepancies highlight how diverse definitions and analytic approaches can contribute to conflicting findings and complicate the interpretation of the evidence in this field. The heterogeneity in this field makes it impossible to present findings in a quantitative meta-analysis, and there is a need for more standardized study designs if studies cannot generate definitive evidence by themselves.

Several recurring methodological issues also emerged. First, most studies were geographically concentrated in Asia and Europe, limiting generalizability to other health care settings and populations, particularly those in South America and Africa. Second, many studies assessed multiple combinations of CGM metrics and CVD outcomes in separate models without prespecified hypotheses or correction for multiple testing. Third, and most importantly, only 3 studies performed prediction modeling analyses, all of which were carried out as secondary analyses. None of the studies reported any external validation or performance metrics beyond discrimination (AUC, sensitivity, and specificity), indicating modest performance [61]. No study applied machine learning methods or used raw CGM time-series data, which may further constrain predictive ability. Thus, the predictive utility of CGM-derived metrics for CVD outcomes remains essentially untested. There is a clear need for sufficiently powered, longitudinal prediction studies using clinical CVD outcomes in ethnically diverse populations [62].

The vast majority of identified studies were cross-sectional, limiting their clinical relevance due to potential reverse causality, as established CVD can alter lifestyle behaviors and glucose patterns. This bias can skew the results in 2 directions.

XSL•FO
RenderX

First, it may produce false-positive associations if distinct CGM patterns only emerge after a CVD event. Second, it can yield false-negative findings if incident CVD, or the resulting intensive medical treatment, masks or attenuates a pre-existing glucose pattern. However, there were too few longitudinal studies to assess if reverse causality systematically skewed the estimates provided by cross-sectional studies and thereby led to divergent results between the 2 study designs.

Collectively, the methodological challenges identified in this review indicate a need for clearer methodological alignment in future studies if systematic reviews are to be feasible. Specifically, researchers should adhere to consensus guidelines, such as the ATTD (Advanced Technologies and Treatments for Diabetes) consensus recommendations [63], together with prespecified covariate adjustment strategies, standardized classification of cardiovascular outcomes, and transparent reporting of analytic decisions. Addressing these areas would improve comparability across studies and strengthen the interpretability of future evidence.

Most studies used CGM data collected specifically for research, with only a few studies drawing on routinely collected real-world data despite the growing prevalence of CGM use. This represents a missed opportunity, as routine data are typically larger, more cost-effective, and more representative of CGM users. Underuse may reflect challenges in accessing data stored on proprietary manufacturer platforms or linking these data to individual health records. Open, publicly available datasets have driven advances in many fields (eg, medical image analysis) [64-66], but no comparable dataset exists for studying CGM data and complications. In the absence of such resources, aligning existing databases with FAIR (findability, accessibility, interoperability, and reusability) principles could help accelerate research in this area [67].

Strengths and Limitations

A key strength of this review was the differentiation between association and prediction studies, highlighting the lack of knowledge on how well CGM-derived metrics perform in CVD prediction models. We performed a more comprehensive literature search, yielding an additional 40 studies compared to a previous review [14]. The distinction between clinical and subclinical CVD allowed a more detailed synthesis of how CGM-derived metrics relate to both CVD manifestation and early vascular changes. Furthermore, we provided a detailed methodological overview and revealed common methodological weaknesses, including variations in the calculation of CGM-derived metrics and the definitions of cardiovascular outcomes.

This review also has limitations. First, the feasibility of synthesizing effect sizes consistently across studies was limited by heterogeneity in study designs, CGM metrics, and CVD outcome definitions. Therefore, this review summarized studies based on *P* values, which is suboptimal, as *P* values are influenced by both the effect size and the sample size [68]. *P* values do not accurately reflect the effect size, clinical relevance, or estimate precision. This greatly limits our ability to compare the strength of associations across studies. Furthermore, underpowered studies are more likely not to find evidence for associations, thereby adding noise to the literature. Second, we reported only the most adjusted models from each study. While this approach was deemed necessary, it may have excluded potentially informative results from alternative model specifications. Third, identifying all relevant studies in this field proved challenging. We decided to limit the search to MEDLINE and Embase only, as these are core databases for biomedical literature searching. Given the resources available to the review team, we were not able to extend the database search further. However, the search retrieved a high number of records, both relevant and irrelevant, owing to inconsistent terminology and overlapping search categories (eg, “blood glucose monitoring” and “glycemic control” both encompass finger-prick measurements). We therefore designed a broad search strategy to ensure that we did not miss any relevant studies in the 2 databases that we chose to search. Acknowledging that searching only 2 databases may have resulted in missing relevant studies, we systematically screened all references and citing articles (backward and forward citation searching) of the included studies. This process resulted in the identification of 2 additional articles [44,69], of which 1 article (Koroleva et al [44]) was not indexed in the medical databases we searched. Nevertheless, to the best of our knowledge, this comprehensive search strategy enabled us to identify more relevant studies than any previous review on this topic.

Conclusion

This scoping review mapped a broad landscape of association studies examining associations between CGM-derived metrics and CVD outcomes, with a smaller number also addressing prediction. The included studies were methodologically heterogeneous, making it difficult to synthesize evidence and draw firm conclusions about clinical cardiovascular risk.

Within these constraints and using statistical significance as a pragmatic indicator of consistency across heterogeneous studies covering different CVD outcomes, TIR was associated with CVD in the largest single study, and MAGE was the CGM-derived metric most consistently associated with CVD outcomes across multiple studies covering subclinical outcomes, coronary artery disease, and severity scores. Notably, MAGE was the only CGM-derived metric to have its predictive value assessed, and it exhibited only modest discriminatory performance. None of the studies used any machine learning–based methods, suggesting that the predictive value of CGM-derived metrics for CVD outcomes and the possibilities of using machine learning–based methods are underexplored. There is a fragmented evidence base in which metric definitions, study designs, and analytical strategies vary widely. In the future, more standardized analytical strategies could enable meta-analyses across individual studies to synthesize more substantial evidence.

XSL•FO
RenderX

Funding

HBT, AAI, BL-J, and AH are employed at Steno Diabetes Center Aarhus, which is funded by a donation from the Novo Nordisk Foundation. AH is supported by a Data Science Emerging Investigator grant (number: NNF22OC0076725) from the Novo Nordisk Foundation. HBT is supported by a research grant from the Danish Diabetes and Endocrine Academy and the Danish Cardiovascular Academy, which are funded by the Novo Nordisk Foundation (grant numbers: NNF22SA0079901 and NNF20SA0067242). THA and ON are employed at the Steno Diabetes Center Copenhagen, a public hospital and research institution under the Capital Region of Denmark, which is partly funded by the Novo Nordisk Foundation. None of the funding bodies had any role in the study.

Data Availability

All data generated or analyzed during this study are included in this article and its supplementary information files.

Authors' Contributions

HBT, BL-J, AH, THA, ON, and AAI conceptualized the study. ON and THA developed the search strategy with feedback from HBT, BL-J, AH, and AAI. ON conducted both searches and the forward and backward citation search. HBT, BL-J, AH, and AAI screened the abstracts. HBT and AAI screened full-text articles. HBT extracted data from the identified studies. AAI verified the data extraction forms. HBT, AH, and AAI analyzed the data and presented the results. HBT wrote the original draft of the manuscript with support from AAI and AH. STA and GF contributed through critical revision of the manuscript, identifying conceptual gaps and helping shape the overall structure and scientific narrative. All authors read, edited, and approved the final version of the manuscript. AH and AAI were responsible for supervising the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string.
[DOCX File, 20 KB - diabetes_v11i1e89374_app1.docx]

Multimedia Appendix 2
Subclinical cardiovascular outcome results.
[DOCX File, 59 KB - diabetes_v11i1e89374_app2.docx]

Multimedia Appendix 3
List of continuous glucose monitoring metrics and cardiovascular disease outcomes found in the literature.
[DOCX File, 23 KB - diabetes_v11i1e89374_app3.docx]

Multimedia Appendix 4
Included studies (with reference details) and their aims.
[XLSX File, 31 KB - diabetes_v11i1e89374_app4.xlsx]

Multimedia Appendix 5
Studies excluded during full-text screening.
[DOCX File, 22 KB - diabetes_v11i1e89374_app5.docx]

Multimedia Appendix 6
Geographical location of the studies.
[DOCX File, 19 KB - diabetes_v11i1e89374_app6.docx]

Multimedia Appendix 7
Overview of the extracted results.
[XLSX File, 35 KB - diabetes_v11i1e89374_app7.xlsx]

Multimedia Appendix 8
List of groupings of adjusted variables.
[DOCX File, 1671 KB - diabetes_v11i1e89374_app8.docx]

XSL•FO
RenderX

Multimedia Appendix 9
Summary of continuous glucose monitoring–derived metrics in each study and cardiovascular disease outcomes.
[DOCX File, 164 KB - diabetes_v11i1e89374_app9.docx]

Multimedia Appendix 10
Full table (with adjustments) of the main findings of the included studies on clinical cardiovascular outcomes.
[DOCX File, 50 KB - diabetes_v11i1e89374_app10.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 127 KB - diabetes_v11i1e89374_app11.pdf]

References

1. Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical update: cardiovascular disease in diabetes mellitus. Circulation 2016 Jun 14;133(24):2459-2502. [doi: 10.1161/CIRCULATIONAHA.116.022194]
2. Kee OT, Harun H, Mustafa N, et al. Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. Cardiovasc Diabetol 2023 Jan 19;22(1):13. [doi: 10.1186/s12933-023-01741-7] [Medline: 36658644]
3. Ndjaboue R, Ngueta G, Rochefort-Brihay C, et al. Prediction models of diabetes complications: a scoping review. J Epidemiol Community Health 2022 Jun 30;76(10):896-904. [doi: 10.1136/jech-2021-217793] [Medline: 35772935]
4. Galbete A, Tamayo I, Librero J, et al. Cardiovascular risk in patients with type 2 diabetes: a systematic review of prediction models. Diabetes Res Clin Pract 2022 Feb;184:109089. [doi: 10.1016/j.diabres.2021.109089] [Medline: 34648890]
5. SCORE2-Diabetes Working Group and the ESC Cardiovascular Risk Collaboration. SCORE2-Diabetes: 10-year cardiovascular risk estimation in type 2 diabetes in Europe. Eur Heart J 2023 Jul 21;44(28):2544-2556. [doi: 10.1093/eurheartj/ehad260] [Medline: 37247330]
6. Vistisen D, Andersen GS, Hansen CS, et al. Prediction of first cardiovascular disease event in type 1 diabetes mellitus: the steno type 1 risk engine. Circulation 2016 Mar 15;133(11):1058-1066. [doi: 10.1161/CIRCULATIONAHA.115.018844] [Medline: 26888765]
7. Zhu T, Li K, Herrero P, Georgiou P. Deep learning for diabetes: a systematic review. IEEE J Biomed Health Inform 2021 Jul;25(7):2744-2757. [doi: 10.1109/JBHI.2020.3040225] [Medline: 33232247]
8. Nimri R, Phillip M, Kovatchev B. Closed-loop and artificial intelligence-based decision support systems. Diabetes Technol Ther 2023 Feb;25(S1):S70-S89. [doi: 10.1089/dia.2023.2505] [Medline: 36802182]
9. Javaid A, Zghyer F, Kim C, et al. Medicine 2032: the future of cardiovascular disease prevention with machine learning and digital health technology. Am J Prev Cardiol 2022 Dec;12:100379. [doi: 10.1016/j.ajpc.2022.100379] [Medline: 36090536]
10. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019 Aug;43(4):383-397. [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
11. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
12. Foreman YD, van Doorn W, Schaper NC, et al. Greater daily glucose variability and lower time in range assessed with continuous glucose monitoring are associated with greater aortic stiffness: the Maastricht Study. Diabetologia 2021 Aug;64(8):1880-1892. [doi: 10.1007/s00125-021-05474-8] [Medline: 33991193]
13. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-412. [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
14. Yapanis M, James S, Craig ME, O'Neal D, Ekinci EI. Complications of diabetes and metrics of glycemic management derived from continuous glucose monitoring. J Clin Endocrinol Metab 2022 May 17;107(6):e2221-e2236. [doi: 10.1210/clinem/dgac034] [Medline: 35094087]
15. Zhang L, Sun XX, Tian QS. Research progress on the association between glycemic variability index derived from CGM and cardiovascular disease complications. Acta Diabetol 2024 Jun;61(6):679-692. [doi: 10.1007/s00592-024-02241-0] [Medline: 38467807]
16. Varga TV, Niss K, Estampador AC, Collin CB, Moseley PL. Association is not prediction: a landscape of confused reporting in diabetes - a systematic review. Diabetes Res Clin Pract 2020 Dec;170:108497. [doi: 10.1016/j.diabres.2020.108497] [Medline: 33068662]
17. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews - JBI Manual for Evidence Synthesis. JBI. 2020. URL: https://jbi-global-wiki.refined.site/space/MANUAL/355862497/10.+Scoping+reviews [accessed 2024-12-15]

XSL•FO
RenderX

18. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Thomsen HB, Isaksen AA, Lebiecka-Johansen B, Andersen TH, Nørgaard O, Hulman A. Continuous glucose monitoring-derived metrics of glycemic control and cardiovascular risk among people with diabetes: a systematic scoping review protocol. Figshare. 2024. URL: https://figshare.com/articles/online_resource/Continuous_glucose_monitoring-derived_metrics_of_glycemic_control_and_cardiovascular_risk_among_people_with_diabetes_a_systematic_scoping_review_protocol/25625445 [accessed 2026-04-20]
20. Helleputte S, Calders P, Rodenbach A, et al. Time-varying parameters of glycemic control and glycation in relation to arterial stiffness in patients with type 1 diabetes. Cardiovasc Diabetol 2022 Dec 9;21(1):277. [doi: 10.1186/s12933-022-01717-z] [Medline: 36494687]
21. Lu J, Wang C, Shen Y, et al. Time in range in relation to all-cause and cardiovascular mortality in patients with type 2 diabetes: a prospective cohort study. Diabetes Care 2021 Feb;44(2):549-555. [doi: 10.2337/dc20-1862] [Medline: 33097560]
22. Martinez M, Santamarina J, Pavesi A, Musso C, Umpierrez GE. Glycemic variability and cardiovascular disease in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2021 Mar;9(1):e002032. [doi: 10.1136/bmjdrc-2020-002032] [Medline: 33762313]
23. Mita T, Katakami N, Okada Y, et al. Continuous glucose monitoring-derived time in range and CV are associated with altered tissue characteristics of the carotid artery wall in people with type 2 diabetes. Diabetologia 2023 Dec;66(12):2356-2367. [doi: 10.1007/s00125-023-06013-3] [Medline: 37750893]
24. Lu J, Ma X, Shen Y, et al. Time in range is associated with carotid intima-media thickness in type 2 diabetes. Diabetes Technol Ther 2020 Feb;22(2):72-78. [doi: 10.1089/dia.2019.0251] [Medline: 31524497]
25. Mo Y, Zhou J, Li M, et al. Glycemic variability is associated with subclinical atherosclerosis in Chinese type 2 diabetic patients. Cardiovasc Diabetol 2013 Jan 15;12(1):15. [doi: 10.1186/1475-2840-12-15] [Medline: 23316958]
26. Wakasugi S, Mita T, Katakami N, et al. Associations between continuous glucose monitoring-derived metrics and arterial stiffness in Japanese patients with type 2 diabetes. Cardiovasc Diabetol 2021 Jan 7;20(1):15. [doi: 10.1186/s12933-020-01194-2] [Medline: 33413339]
27. Magri CJ, Mintoff D, Camilleri L, Xuereb RG, Galea J, Fava S. Relationship of hyperglycaemia, hypoglycaemia, and glucose variability to atherosclerotic disease in type 2 diabetes. J Diabetes Res 2018;2018:7464320. [doi: 10.1155/2018/7464320] [Medline: 30140707]
28. Su G, Mi S, Tao H, et al. Association of glycemic variability and the presence and severity of coronary artery disease in patients with type 2 diabetes. Cardiovasc Diabetol 2011 Feb 25;10(1):19. [doi: 10.1186/1475-2840-10-19] [Medline: 21349201]
29. Tang X, Li S, Wang Y, et al. Glycemic variability evaluated by continuous glucose monitoring system is associated with the 10-y cardiovascular risk of diabetic patients with well-controlled HbA1c. Clin Chim Acta 2016 Oct 1;461:146-150. [doi: 10.1016/j.cca.2016.08.004] [Medline: 27502250]
30. Zhang X, Xu X, Jiao X, Wu J, Zhou S, Lv X. The effects of glucose fluctuation on the severity of coronary artery disease in type 2 diabetes mellitus. J Diabetes Res 2013;2013:576916. [doi: 10.1155/2013/576916] [Medline: 23936867]
31. Haddaway NR, Grainger MJ, Gray CT. Citationchaser: a tool for transparent and efficient forward and backward citation chasing in systematic searching. Res Synth Methods 2022 Jul;13(4):533-545. [doi: 10.1002/jrsm.1563] [Medline: 35472127]
32. EPPI Reviewer. URL: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914 [accessed 2026-04-28]
33. Georeli E, Stamati A, Dimitriadou M, et al. Assessment of arterial stiffness in paediatric patients with type 1 diabetes mellitus. J Diabetes Complicat 2024 Aug;38(8):108782. [doi: 10.1016/j.jdiacomp.2024.108782]
34. Pulkkinen MA, Tuomaala AK, Hero M, Gordin D, Sarkola T. Motivational Interview to improve vascular health in adolescents with poorly controlled type 1 diabetes (MIAD): a randomized controlled trial. BMJ Open Diabetes Res Care 2020 Jul;8(1):e001216. [doi: 10.1136/bmjdrc-2020-001216] [Medline: 32723754]
35. Hoffman RP, Dye AS, Huang H, Bauer JA. Effects of glucose control and variability on endothelial function and repair in adolescents with type 1 diabetes. ISRN Endocrinol 2013;2013:876547. [doi: 10.1155/2013/876547] [Medline: 24490081]
36. Morandi A, Piona C, Corradi M, et al. Risk factors for pre-clinical atherosclerosis in adolescents with type 1 diabetes. Diabetes Res Clin Pract 2023 Apr;198:110618. [doi: 10.1016/j.diabres.2023.110618] [Medline: 36906234]
37. Piona C, Marigliano M, Mancioppi V, et al. Glycemic variability and time in range are associated with the risk of overweight and high LDL-cholesterol in children and youths with type 1 diabetes. Horm Res Paediatr 2025;98(1):59-65. [doi: 10.1159/000535554] [Medline: 38104553]
38. Bergdahl E, Forsander G, Sundberg F, Milkovic L, Dangardt F. Investigating the presence and detectability of structural peripheral arterial changes in children with well-regulated type 1 diabetes versus healthy controls using ultra-high frequency ultrasound: a single-centre cross-sectional and case-control study. EClinicalMedicine 2025 Mar;81:103097. [doi: 10.1016/j.eclinm.2025.103097] [Medline: 40034566]
39. Peña AS, Couper JJ, Harrington J, et al. Hypoglycemia, but not glucose variability, relates to vascular function in children with type 1 diabetes. Diabetes Technol Ther 2012 Jun;14(6):457-462. [doi: 10.1089/dia.2011.0229] [Medline: 22313018]

XSL•FO
RenderX

40. Chen Y, Jia T, Yan X, Dai L. Blood glucose fluctuations in patients with coronary heart disease and diabetes mellitus correlates with heart rate variability: a retrospective analysis of 210 cases. Niger J Clin Pract 2020 Sep;23(9):1194-1200. [doi: 10.4103/njcp.njcp_529_19] [Medline: 32913156]
41. Gordin D, Rönnback M, Forsblom C, Mäkinen V, Saraheimo M, Groop PH. Glucose variability, blood pressure and arterial stiffness in type 1 diabetes. Diabetes Res Clin Pract 2008 Jun;80(3):e4-e7. [doi: 10.1016/j.diabres.2008.01.010] [Medline: 18325620]
42. Borg R, Kuenen JC, Carstensen B, et al. HbA_1 (c) and mean blood glucose show stronger associations with cardiovascular disease risk factors than do postprandial glycaemia or glucose variability in persons with diabetes: the A1C-Derived Average Glucose (ADAG) study. Diabetologia 2011 Jan;54(1):69-72. [doi: 10.1007/s00125-010-1918-2] [Medline: 20886203]
43. Shu-Hua M, Gong S, Zhao L, et al. Comparison of glycemic variability and glycated hemoglobin as risk factors of coronary artery disease in patients with undiagnosed diabetes. Chin Med J 2012;125(1):38-43. [doi: 10.3760/cma.j.issn.0366-6999.2012.01.008]
44. Koroleva EA, Korbut AI, Bulumbaeva DM, et al. Modeling of risk factors for peripheral artery disease in patients with type 2 diabetes. Presented at: IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON); Nov 11-13, 2022. [doi: 10.1109/SIBIRCON56155.2022.10017018]
45. He R, Xuan Y, Zhu L, et al. Low blood glucose index associated with cardiovascular events in diabetic hemodialysis patients. Blood Purif 2023;52(9-10):824-834. [doi: 10.1159/000531964] [Medline: 37607516]
46. Wei W, Zhao S, Fu SL, et al. The association of hypoglycemia assessed by continuous glucose monitoring with cardiovascular outcomes and mortality in patients with type 2 diabetes. Front Endocrinol (Lausanne) 2019;10:536. [doi: 10.3389/fendo.2019.00536] [Medline: 31447782]
47. Bezerra MF, Neves C, Neves JS, Carvalho D. Time in range and complications of diabetes: a cross-sectional analysis of patients with type 1 diabetes. Diabetol Metab Syndr 2023 Nov 27;15(1):244. [doi: 10.1186/s13098-023-01219-2] [Medline: 38008747]
48. De Meulemeester J, Charleer S, Visser MM, De Block C, Mathieu C, Gillard P. The association of chronic complications with time in tight range and time in range in people with type 1 diabetes: a retrospective cross-sectional real-world study. Diabetologia 2024 Aug;67(8):1527-1535. [doi: 10.1007/s00125-024-06171-y] [Medline: 38787436]
49. Deng Z, Wang S, Lu J, et al. Interaction between haptoglobin genotype and glycemic variability on diabetic macroangiopathy: a population-based cross-sectional study. Endocrine 2023 Nov;82(2):311-318. [doi: 10.1007/s12020-023-03484-7] [Medline: 37615814]
50. El Malahi A, Van Elsen M, Charleer S, et al. Relationship between time in range, glycemic variability, HbA1c, and complications in adults with type 1 diabetes mellitus. J Clin Endocrinol Metab 2022 Jan 18;107(2):e570-e581. [doi: 10.1210/clinem/dgab688] [Medline: 34534297]
51. Guo J, Wang J, Zhao Z, Yu L. Association between glycemic control assessed by continuous glucose monitoring and stroke in patients with atrial fibrillation and diabetes mellitus. Ann Palliat Med 2021 Aug;10(8):9157-9164. [doi: 10.21037/apm-21-2198] [Medline: 34488401]
52. Li K, Liu C, Zhu T, Herrero P, Georgiou P. GluNet: a deep learning framework for accurate glucose forecasting. IEEE J Biomed Health Inform 2020 Feb;24(2):414-423. [doi: 10.1109/JBHI.2019.2931842] [Medline: 31369390]
53. Sheng X, Li T, Hu Y, Xiong CS, Hu L. Correlation between blood glucose indexes generated by the flash glucose monitoring system and diabetic vascular complications. Diabetes Metab Syndr Obes 2023;16:2447-2456. [doi: 10.2147/DMSO.S418224] [Medline: 37608851]
54. Watanabe M, Kawai Y, Kitayama M, et al. Diurnal glycemic fluctuation is associated with severity of coronary artery disease in prediabetic patients: possible role of nitrotyrosine and glyceraldehyde-derived advanced glycation end products. J Cardiol 2017 Apr;69(4):625-631. [doi: 10.1016/j.jjcc.2016.07.001] [Medline: 27470137]
55. Gensini GG. A more meaningful scoring system for determining the severity of coronary heart disease. Am J Cardiol 1983 Feb;51(3):606. [doi: 10.1016/s0002-9149(83)80105-2] [Medline: 6823874]
56. SYNTAX Score. URL: https://syntaxscore.org/index.php [accessed 2025-05-28]
57. Kovatchev B, Cobelli C. Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes. Diabetes Care 2016 Apr;39(4):502-510. [doi: 10.2337/dc15-2035] [Medline: 27208366]
58. Castaldo E, Sabato D, Lauro D, Sesti G, Marini MA. Hypoglycemia assessed by continuous glucose monitoring is associated with preclinical atherosclerosis in individuals with impaired glucose tolerance. PLoS ONE 2011;6(12):e28312. [doi: 10.1371/journal.pone.0028312] [Medline: 22164268]
59. Mesa A, Giménez M, Pueyo I, et al. Hyperglycemia and hypoglycemia exposure are differentially associated with micro- and macrovascular complications in adults with type 1 diabetes. Diabetes Res Clin Pract 2022 Jul;189:109938. [doi: 10.1016/j.diabres.2022.109938] [Medline: 35662616]
60. Johnston SS, Conner C, Aagren M, Smith DM, Bouchard J, Brett J. Evidence linking hypoglycemic events to an increased risk of acute cardiovascular events in patients with type 2 diabetes. Diabetes Care 2011 May;34(5):1164-1170. [doi: 10.2337/dc10-1915] [Medline: 21421802]

XSL•FO
RenderX

61. Van Calster B, Collins GS, Vickers AJ, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. Lancet Digit Health 2025 Dec;7(12):100916. [doi: 10.1016/j.landig.2025.100916] [Medline: 41391983]
62. Ajjan RA, Battelino T, Seufert J, et al. Do continuous glucose monitoring (CGM) metrics predict macrovascular and microvascular complications in diabetes? The FACULTY protocol of a retrospective real-world cohort study. BMJ Open 2025 Jan 8;15(1):e085961. [doi: 10.1136/bmjopen-2024-085961] [Medline: 39779259]
63. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. Diabetes Care 2017 Dec;40(12):1631-1640. [doi: 10.2337/dc17-1600] [Medline: 29162583]
64. NIH chest x-rays. Kaggle. URL: https://www.kaggle.com/datasets/nih-chest-xrays/data [accessed 2025-07-16]
65. CheXpert. Stanford ML Group. URL: https://stanfordmlgroup.github.io/competitions/chexpert/ [accessed 2025-07-16]
66. APTOS 2019 blindness detection. Kaggle. URL: https://www.kaggle.com/competitions/aptos2019-blindness-detection [accessed 2025-07-16]
67. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018. [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
68. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature New Biol 2019 Mar;567(7748):305-307. [doi: 10.1038/d41586-019-00857-9] [Medline: 30894741]
69. Cai L, Shen W, Li J, et al. Association between glycemia risk index and arterial stiffness in type 2 diabetes. J of Diabetes Invest 2024 May;15(5):614-622. [doi: 10.1111/jdi.14153]

Abbreviations

AUC: area under the receiver operating characteristic curve
CGM: continuous glucose monitoring
CV: coefficient of variation
CVD: cardiovascular disease
HbA_{1c}: hemoglobin A_{1c}
MAGE : mean amplitude of glycemic excursions
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TAR: time above range
TBR : time below range
TIR: time in range
TITR: time in tight range

Edited by S Li; submitted 11.Dec.2025; peer-reviewed by J Li, Y Pan; revised version received 27.Feb.2026; accepted 08.Mar.2026; published 06.May.2026.

Please cite as:
Thomsen HB, Lebiecka-Johansen B, Nørgaard O, Helms Andersen T, Andersen ST, Fagherazzi G, Hulman A, Isaksen AA
Continuous Glucose Monitoring–Derived Metrics and Cardiovascular Risk Among People With Diabetes: Systematic Scoping Review
JMIR Diabetes 2026;11:e89374
URL: https://diabetes.jmir.org/2026/1/e89374
doi:10.2196/89374

© Helene Bei Thomsen, Benjamin Lebiecka-Johansen, Ole Nørgaard, Tue Helms, Signe Toft Andersen, Guy Fagherazzi, Adam Hulman, Anders Aasted Isaksen. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 6.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Quantitative Framework for Evaluating the Performance of Algorithm-Directed Whole-Population Remote Patient Monitoring: Tutorial for Type 1 Diabetes Care

Jamie Kurtzig[1]; Ananta Addala[1,2], DO, MPH; Franziska K Bishop[1], MS, CDCES; Paul Dupenloup[3], MS; Johannes O Ferstad[3], PhD; Ramesh Johari[3], PhD; David M Maahs[1,2], MD, PhD; Priya Prahalad[1,2], MD, PhD; Dessi P Zaharieva[1,2], PhD, CDCES; David Scheinker[1,2,3,4,5], PhD

[1]Division of Endocrinology, Department of Pediatrics, Stanford Medicine, 291 Campus Drive, Stanford, CA, United States
[2]Stanford Diabetes Research Center, Stanford University, Stanford, CA, United States
[3]Department of Management Science and Engineering, Stanford University, Stanford, CA, United States
[4]Clinical Excellence Research Center, Stanford University, Stanford, CA, United States
[5]Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
David Scheinker, PhD
Division of Endocrinology, Department of Pediatrics, Stanford Medicine, 291 Campus Drive, Stanford, CA, United States

Abstract

Clinics continue to adopt care models shaped by the algorithmic analysis of continuous glucose monitoring (CGM) data, such as remote patient monitoring for type 1 diabetes. No clinic-facing quantitative framework currently exists to track the impact of such algorithm-directed care on patient outcomes and clinical workload. We used CGM data from the Teamwork, Targets, Technology, and Tight Control (4T) Study (Pilot n=135 and Study 1 n=133), in which algorithms enable precision, whole-population care by directing clinician attention to patients with deteriorating glucose management. Youth meeting criteria for clinical review are then contacted by Certified Diabetes Care and Education Specialists. Through iterative data analysis and meetings with a variety of stakeholders, we identified metrics for reviewing and revising clinical workloads, glucose management, and timeliness of care. For each metric, we developed an interactive dashboard to provide clinical and administrative leaders with an overview of the program. The metrics to track clinical workload were the total number of youths (1) in the program, (2) in each study, and (3) cared for by each clinician. The metrics to track glucose management were the number of youths meeting each criterion for review, including (4) total, (5) for each clinician, and (6) for each study. The metric to track timeliness of care was (7) the number of days since meeting criteria for clinical review. When presented at regular program leadership meetings, the metrics facilitated data-driven decision-making about clinical and operational components of the program. In this paper, we describe the process of developing and operationalizing this reproducible, clinician-facing key performance indicator tool to monitor an algorithm-enabled remote patient monitoring program. As the role of algorithms grows in directing clinical effort and prioritizing patients for care, this framework may help clinics track clinical workload, patient outcomes, and the timeliness of care.

Trial Registration: Clinicaltrials.gov NCT03968055; https://clinicaltrials.gov/study/NCT03968055 and Clinicaltrials.gov NCT04336969; https://clinicaltrials.gov/study/NCT04336969

(JMIR Diabetes 2026;11:e72676) doi:10.2196/72676

KEYWORDS

remote patient monitoring; continuous glucose monitoring; type 1 diabetes; chronic disease; precision medicine; quality improvement

Introduction

Remote patient monitoring (RPM) programs for type 1 diabetes (T1D) care often use data provided by continuous glucose monitoring (CGM) systems [1,2]. RPM for T1D is shown to improve hemoglobin A_{1c} (HbA_{1c}), quality of life, and glucose metrics in a variety of health care settings [3-5]. Numerous centers report the use of RPM for T1D care, and many include some level of algorithmic support in identifying patients for prioritization and clinical review or contact [6-10]. The growing use of CGM drives the demand for algorithms to manage data and inform care delivery, increasing the complexity of clinic operations [11,12]. Methods have been proposed for the retrospective analysis of the potential impact of changes to an algorithm that supports patient care, but there are few prospective approaches to monitoring the impact of algorithm-directed care [13].

XSL•FO
RenderX

Numerous studies report the use of an algorithm to inform patient care [6,10]. At the University of California, Davis, a single-center pilot study evaluated RPM for patients with newly diagnosed pediatric T1D [6]. The principal investigator reviewed each patient's glucose data daily using the electronic health record (EHR) and Tidepool Data Platform (Tidepool Project). Monitoring began at hospital discharge and continued until the patient's first outpatient clinic appointment. Patients were contacted as needed to adjust insulin doses based on their data. A statewide study in South Carolina provided RPM to underserved populations [10]. Research staff regularly reviewed patients' glucose data and alerted the appropriate clinic if glucose values were abnormal. Research staff also generated regular reports identifying patients who either failed to transmit their glucose data or had high HbA_{1c} values. These reports were sent to appropriate local clinical teams, who were encouraged, but not required, to adjust patient treatment based on the data.

These, and other RPM programs, share common features that drive care provider workload, patient outcomes, and the cadence with which care providers engage with patients. Those measures are examples of key performance indicators (KPIs). KPIs track and can drive an organization's progress toward strategic goals [14]. The potential of KPIs to monitor these aspects of care has been examined, but not in the context of RPM [14]. We sought to extend earlier findings and principles of the use of KPIs in health care to the setting of RPM.

In the Teamwork, Targets, Technology, and Tight Control (4T) Study 1 cohort, the mean HbA_{1c} was 6.58% at 12 months after diagnosis [9]. The Stanford 4T Study whole-population RPM tool, Timely Interventions for Diabetes Excellence (TIDE; Stanford researchers), uses CGM data to prioritize patients with T1D for personalized review and contact by a Certified Diabetes Care and Education Specialist (CDCES), a health care professional who educates and supports people with diabetes to improve health and reduce the risk of complications [15]. The TIDE algorithm informs glucose management, how patients are treated, and the regularity with which patients receive CDCES messages. The 4T Study developed a protocol within which CDCESs have safely and effectively adjusted insulin doses and increased patient engagement [16]. Within the 4T Study, CDCESs review patient data via the TIDE dashboard. They then interpret the algorithm-directed prioritization and patient data to recommend insulin dose adjustment and contact the patient via the EHR [16]. The 4T Study and the use of TIDE have been associated with significant, equitable improvements in glycemic outcomes, as well as reduced CDCES workload and burden [2,9,12,17].

To our knowledge, no clinician-facing quantitative framework is available to track how algorithm-directed care impacts clinical workload, patient glucose management, and timeliness of care. Such quantitative frameworks may be helpful for clinics that remotely access patient data, provide RPM-based care, and use algorithms to direct care delivery by, for example, identifying or prioritizing patients requiring care. This work aims to describe the process of developing and implementing KPIs and corresponding metrics to help clinicians, researchers, and administrators monitor clinical workload, patient outcomes, and timeliness of care.

Methods

The 4T Study

This work was conducted within the 4T Study. The 4T Study aims to improve care for newly diagnosed youth with T1D in Stanford's Pediatric Endocrinology clinics. A summary of the 4T program is provided below, and details of the program have been published [9].

The 4T Study includes the 4T Pilot Study and 4T Study 1, as well as long-term follow-ups for each study. The long-term follow-up programs assess the ongoing impact of the 4T Pilot and Study 1 over an extended period. The 4T Pilot Study (n=135) enrolled participants from July 2018 to June 2020 [7,18]. 4T Pilot participants diagnosed before March 2019 were offered CGM, but RPM was not yet available. Those diagnosed between March 2019 and January 2020 were offered non–algorithm-enabled RPM (eg, Dexcom Clarity), which provided individual T1D insights rather than population-wide overviews such as TIDE. Pilot participants diagnosed during and after January 2020 were enrolled in the algorithm-enabled care platform for RPM. Participants diagnosed before January 2020 could transition to algorithm-enabled care when it launched. 4T Study 1 (n=133) enrolled participants from June 2020 to March 2022 and focused on initiating CGM use within the first 30 days after T1D diagnosis.

Step 1: Identify an Appropriate Algorithm-Enabled Care Model

The quantitative RPM framework presented here is designed to support an algorithm-enabled care model. The 4T Study uses TIDE to support asynchronous review of CGM data for enrolled participants. All 4T participants had the choice to opt into TIDE when it launched.

Participants uploaded their CGM data via a personal smart device or study-provided iPod Touch (Apple Inc) to Dexcom Clarity. TIDE automatically pulled CGM data from the 4T Study's Dexcom Clarity clinic account using a Python script, analyzed the data, and displayed results in an interactive data visualization dashboard created in Tableau (Salesforce). Guided by CGM clinical consensus metrics, the TIDE algorithm ranked participants based on the urgency of data review and contact by a CDCES [2,19,20].

Each week, the TIDE platform tracked whether participants met the following clinical categories: spending less than 65% of CGM time in range (TIR; 70 - 180 mg/dL), more than 4% time below range (TBR) level 1 (<70 mg/dL), more than 1% TBR level 2 (<55 mg/dL), a drop in TIR of more than 15% points from the previous week, more than 50% missing CGM data, or meeting clinical targets [20]. Meeting targets refers to all participants who do not meet any of the other criteria (ie, they have >65% TIR, <4% TBR level 1, <1% TBR level 2, <15% point drop in TIR, and <50% missing CGM data). These clinical categories are based on clinical consensus guidelines [20]. Some metrics were updated between the Pilot Study and Study 1; for

XSL•FO
RenderX

example, the TIR target was changed from 70% to 65%, changing the criteria flagged.

The CDCES team members used TIDE to view a list of participants, ranked by their clinical risk likelihood as defined by the CGM metrics listed above. CDCES team members then took appropriate follow-up steps if needed, such as sending secure messages with care recommendations to participants through the EHR. In both the 4T Pilot and Study 1, participant data were initially reviewed weekly for the first year after T1D diagnosis and monthly thereafter [9,12]. After the conclusion of each study, participants had the option to participate in an ongoing long-term follow-up program with monthly data reviews conducted by a CDCES.

Step 2: Choose Key Performance Indicators

A metric is a quantitative measure of an aspect of a care process or the patient population. A KPI combines metrics into a high-level measure or visualization that reflects strategic goals and provides insights into the overall success and performance of an organization [14]. Commonly reported KPIs for health care organizations include patient waiting time, patient health status, patient quality of life, patient safety, and clinical efficiency [14].

Potential KPIs were eligible for inclusion if they aligned with the aims of the 4T Study and could be operationalized using patient CGM data within the existing 4T and TIDE infrastructure.

Using the above evidence-based list of recommended KPIs, inclusion criteria, and input from 4T clinicians, we selected clinical workload, glucose management, and timeliness of care for inclusion in the KPI framework.

We did not include safety or quality-of-life KPIs, as these are the focus of dedicated retrospective analyses (rather than ongoing KPIs). The 4T Study improved HbA_{1c}, TIR, and patient-reported outcomes (PROs) with minimal TBR (<70 mg/dL) as published [7,9,17,21,22]. The algorithm does not recommend any specific insulin dose adjustment (that is done by the CDCES team) but rather alerts the CDCES that the CGM glucose data require attention.

Step 3: Propose Initial Metrics for Each Key Performance Indicator

In 2022, the engineering team began developing a standardized approach to monitor the clinical workload, patient glucose management, and timeliness of care associated with the use of TIDE. Numerous ad hoc metrics were developed and measured based on the experiences of the 4T Study team, CDCES team, and administrative team [2].

The metrics were calculated from data continuously pulled from servers on each participant's data completeness, the study cohort they were enrolled in, the most recent week they were shown in the TIDE dashboard, their CGM data, and the CDCES responsible for reviewing their data.

Step 4: Develop Visualizations for Each Metric

We developed an instantiation of the quantitative framework with clinic-specific metrics deployed in an interactive dashboard engineered to streamline the monitoring and visualization of the entire program. All visualizations developed in Tableau were double-checked with manual calculations. We note that participants were enrolled in and departed from the 4T Study continuously; thus, the number of participants shown at any time in the visualizations is less than the total study population.

Step 5: Iterate and Obtain Feedback

Feedback Sessions

At regular TIDE meetings, engineers presented the KPI framework and requested feedback. These meetings evaluated the perceived utility of this framework by collecting qualitative insights into the team's firsthand experiences, perceptions, and suggestions regarding the framework's implementation and impact. The meetings represented a diverse range of perspectives, currently including 5 CDCESs, 4 physicians, 4 engineers, 4 biostatisticians, 3 clinical research coordinators, as well as a staff scientist, exercise physiologist, postdoctoral fellow, and psychologist. All team members were encouraged to provide feedback.

Based on the feedback, inappropriate metrics were discarded, new metrics were suggested, visualizations were selected, chosen metrics and visualizations were refined, and actions were taken based on insights from the framework. Feedback was implemented when the team reached a consensus on a metric, visualization, or decision.

There were 10 TIDE meetings where engineers received feedback on the KPI framework. These semistructured meetings each lasted approximately 15 minutes and were conducted via Zoom (Zoom Video Communications). Questions followed a standardized interview guide (Section 1 in Multimedia Appendix 1) and focused on metric interpretability, usefulness, workflow fit, and decision-making impact. The engineering team took detailed notes (excluding any protected patient information) and sent the notes back to meeting participants for confirmation. We synthesized and grouped feedback by metric and visualization, noting discussion points, decisions, action items, and new ideas.

Based on the TIDE meetings, 2 smaller, more specific groups were directly interviewed to expand on and ensure feedback. These as-needed interviews lasted 30 minutes to an hour and were conducted via Zoom. Questions followed the same interview guide as the TIDE meetings (Section 1 in Multimedia Appendix 1), but more emphasis was placed on the specific team's or individual's expertise.

Select Metrics and Visualizations

Metrics and visualizations were discarded based on consensus from TIDE meetings. Metrics and visualizations were retained if (1) high and low values were readily understood by physicians and CDCESs, (2) they supported at least one clinical or operational decision, (3) they could be automatically computed using existing data and workflows, and (4) they provided unique information not given by any other retained metric. Metrics and visualizations were discarded if they failed one or more criteria. The clinical team assessed interpretability and actionability,

XSL•FO
RenderX

and the engineering team evaluated feasibility. Everyone could comment on redundancy.

Throughout our process, 6 metric categories were discarded (Table 1). The first category of discarded metrics included measures of the number of CDCES messages and contacts through the EHR, which included metrics such as overall message count, contact count, TIR by message count, CGM time worn by message count, patients contacted versus suggested for contact, and patient response rate to clinician messages. These message and contact-related metrics were discarded due to a mismatch with our current workflow. CDCES messages are written and received within the EHR (Epic), while the KPI dashboard collects data from the TIDE dashboard. TIDE does not integrate or connect with the EHR. Additionally, CDCESs may not be able to act on insights from these metrics (ie, sending more messages or following up more) because of their limited time and availability.

Table . Excluded metrics and why they were excluded.

Metric	Reasons for exclusion
Number of CDCES[a] electronic health record messages and contacts per participant	• Does not integrate with current workflow • May not drive decisions due to limited clinician time
Diabetes duration stratified by clinical categories and TIR[b]	• Does not drive clinical decisions
Average glucose stratified by CDCES and study (or population of interest)	• Does not provide any additional insights or drive additional decisions
Patient and family satisfaction ratings over time	• Does not integrate with current workflow • 4T Study has already retrospectively investigated this
Patient safety: number of adverse events by CDCES and study (or population of interest)	• 4T Study has already retrospectively investigated this
Clinical workload: CDCES time spent on RPM	• Does not integrate with current workflow

[a]CDCES: Certified Diabetes Care and Education Specialist.

[b]TIR: time in range.

Second were measures of how diabetes duration may impact the clinical categories and CGM metrics, as well as the average diabetes duration for each of the CDCES's patients. These were ultimately excluded since they would not drive day-to-day clinical or operational decisions and would be more appropriate for investigation in retrospective analyses.

Third were measures of participants' average glucose. These were excluded since TIR is highly correlated and considered a more comprehensive measure [20,23].

Fourth, we did not include patient and family satisfaction metrics in the current KPI framework due to the lack of workflow integration. Currently, PROs are collected using REDCap (Research Electronic Data Capture)-based surveys, while the KPI dashboard uses information from the TIDE dashboard. Additionally, several 4T papers have already demonstrated positive PROs [8,22,24,25].

Fifth, we did not include safety measures because previous publications have shown that the 4T Study is associated with lower HbA_{1c} values, improved TIR, minimal TBR, and no unexpected serious adverse events [7,9,17,18].

Finally, we did not include an objective measure of CDCES time in the TIDE dashboard. Prior self-reported data from CDCES indicate approximately 15 minutes per complete patient review and an approximately 60% drop in time per patient per week with TIDE use (from 3.2 to 1.3 minutes) [2,26]. However, because TIDE's host platform (Tableau) lacks session-duration logging, and CDCESs juggle other tasks, objective timing does not yet fit into the workflow. In the interim, the number of patients cared for by each CDCES is a useful proxy of workload. We may supplement the current proxy with CDCES time spent in the Epic EHR. Epic Signal logs task-specific activity (eg, chart review and patient messages), and these data can feed into the KPI dashboard via an application programming interface.

Refine Chosen Metrics and Visualizations

The next stage was an analytical approach to refine the chosen metrics based on an analysis of the data generated and the structured TIDE meetings with the research and clinical staff. In meetings, the stated goals of the process were to establish a more robust and comprehensive set of metrics for monitoring clinical workload, patient glucose management, and timeliness of care in the presence of algorithm-directed care. Interview questions and additional details are provided in Section 1 in Multimedia Appendix 1.

The above process of developing, evaluating, and soliciting feedback (steps 2 - 5) was used iteratively to develop and refine KPIs and metrics based on these KPIs.

To monitor clinical workload in the 4T Study, the following metrics were identified: (1) the number of youths in the 4T program, (2) in each study, and (3) cared for by each CDCES. For tracking participant glucose management, the metrics include the number of youths meeting each criterion for clinical review, both (4) overall and for each (5) CDCES and (6) study. Initially, the metric for tracking the timeliness of care was the number of days since each participant's last day meeting criteria for clinical review. This was later revised to (7) include only participants who have not been seen for more than 20 days and rank in the top 25% of patients with the longest time since

XSL•FO
RenderX

meeting criteria. The detailed definitions of each metric are provided in Section 2 in Multimedia Appendix 1.

To visualize each included metric related to clinical workload, we used different colors to represent the number of patients requiring review, meeting targets, and missing >50% CGM data. For each of those 3 criteria, this was reported as the number of distinct patient IDs.

In the visualization of metrics related to patient outcomes, we used different colors to represent the number of patients in each clinical category. The clinical categories are less than 65% TIR, more than 4% TBR level 1, more than 1% TBR level 2, a drop of more than 15 percentage points in TIR, and less than 50% CGM wear time. For each clinical category, this is defined as the number of distinct patient IDs of participants flagged.

For the visualization of the metric related to timeliness of care, we displayed participant information only if more than 20 days had passed since the participant was last shown in the TIDE dashboard, and the participant was in the highest quartile for delayed care. The threshold of 20 days was selected based on a retrospective analysis of historical TIDE data, which revealed that the 75th percentile of time between participant appearances in the TIDE dashboard was 20 days.

For all metrics and visualizations presented, gaps or valleys in the data occur when, primarily due to the cadence of reviews, fewer CDCESs used TIDE during that period.

Step 6: Use the Metrics to Drive Decisions

Feedback from all care team members highlighted 7 metrics as essential for monitoring clinical workload, patient glucose management, and timeliness of care (Table 2). Each metric is associated with a KPI displayed in an interactive figure with various filters and categorizations (Table 2).

Table . Included metrics, the corresponding key performance indicators, and the decisions the metrics drive.

Metric	Corresponding KPI[a]	Decisions the metric drives
Number of participants in the RPM[b] program	Clinical workload	• Modifying recruitment efforts or stringency of the criteria for participant review
Number of participants reviewed by each CDCES[c] in the program	Clinical workload	• Redistributing participants between the CDCESs
Number of participants in the RPM program for each study (or population of interest)	Clinical workload	• Modifying recruitment efforts
Number of participants in the RPM program per clinical category	Patient glucose management	• Hosting a session for CDCESs to teach their patients how to prevent low and high blood sugars or improve CGM[d] connectivity
Number of participants in the RPM program per clinical category and reviewing CDCES	Patient glucose management	• Evaluating the effectiveness of a CDCES's new initiative
Number of participants in the RPM program per clinical category and study (or population of interest)	Patient glucose management	• Identifying health disparities • Changing the review frequency of a specific participant population
Number of days since participants' most recent date of being displayed in the RPM program	Timeliness of care	• Contacting a participant that has not been flagged for review recently

[a]KPI: key performance indicator.

[b]RPM: remote patient monitoring.

[c]CDCES: Certified Diabetes Care and Education Specialist.

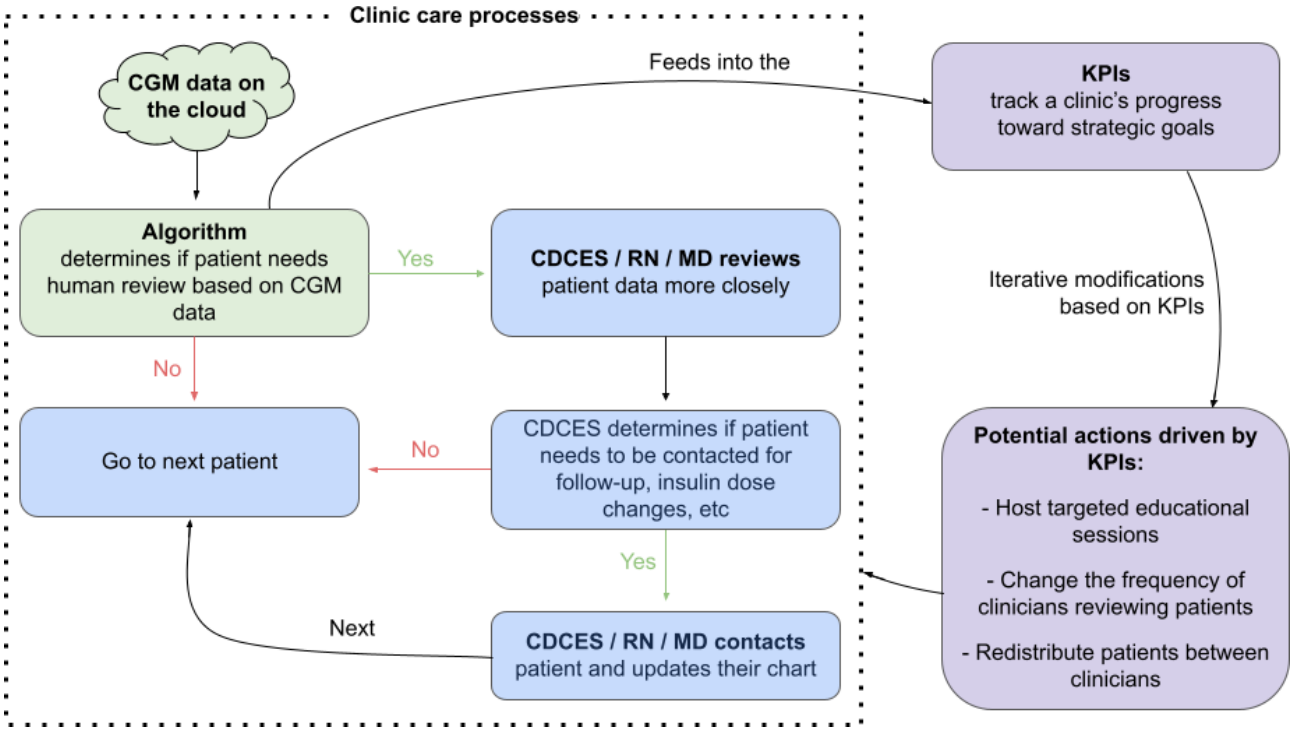
[d]CGM: continuous glucose monitoring.

At regular TIDE team meetings with clinical and administrative 4T team members, the dashboard was presented, and the metrics were discussed to facilitate an overview of the program. The KPIs were developed and used in parallel with the algorithm-enabled clinical care model and any related clinical studies. While the clinic provides patient care using the algorithm, the KPIs are based on the data gathered, are reviewed at a cadence set by the clinic, and may be used to adjust the care model or workflows (Figure 1). For key insights informed by each metric, refer to Section 3 in Multimedia Appendix 1.

Figure 1 shows how clinical care processes and the KPI framework interact. As noted in Step 1, the RPM algorithm ranks patients based on how urgently they need care. A CDCES then reviews patient CGM data and messages them to provide recommendations and education. As discussed in Step 3, the KPI dashboard pulls information from the RPM algorithm. The KPI framework's metrics and associated visualizations help inform clinical and operational decisions and actions that modify the clinical care process.

XSL•FO
RenderX

Figure 1. Care model and framework. CDCES: Certified Diabetes Care and Education Specialist; CGM: continuous glucose monitoring; KPI: key performance indicator; MD: medical doctor; RN: registered nurse.



Ethical Considerations

The study was approved by Stanford's Institutional Review Board (ClinicalTrials.gov: NCT03968055 and NCT04336969; IRB Protocol 52812). Patients and families who consented were enrolled in the 4T Study. Those who did not consent received standard clinical care, which allowed early CGM initiation but did not include RPM. Participants aged 18 years or older or legal guardians of minors provided informed consent, while youth aged 7 years or older provided assent prior to study initiation. Participants in the 4T Study were between the ages of 6 months and 21 years and started on CGM (Dexcom G6; Dexcom Inc) within the first month of T1D diagnosis. Youth and families were not reimbursed for their participation in TIDE, but they could receive compensation for completing PRO surveys, doing home HbA_{1c} fingerstick tests, participating in focus groups, and participating in the 4T exercise substudy. 4T participants also received a small monetary gift (US $10) for their birthdays as a thank you for their continued participation.

Results

Participants

Of the 268 participants in the 4T Pilot and Study 1, a total of 222 were eligible for RPM and were included in the KPI framework (Table 3). All 4T Study 1 participants (n=133) were enrolled in TIDE and included in the KPI framework. In the Pilot Study, participants diagnosed between July 2018 and February 2019 (n=46) were not enrolled in TIDE or included in the KPI framework, and the other 89 were included [7]. The 4T Pilot Study and 4T Study 1 enrolled 92% and 84%, respectively, of all newly diagnosed patients seen in our clinic.

XSL•FO
RenderX

Table . Participant demographics for the 4T Pilot and 4T Study 1.

Characteristics	4T Pilot	4T Study 1
Study length	2018-2020	2020-2022
Participants, n (%)	135 (50.4)	133 (49.6)
Participants on TIDE[a], n (%)	89 (65.9)	133 (100)
Age (years), median (IQR)	10 (7-13)	11 (6-14)
Gender, n (%)		
Women	64 (47.4)	59 (44.4)
Men	71 (52.6)	74 (55.6)
Self-identified race and ethnicity, n (%)		
Non-Hispanic White	53 (39.3)	52 (39.1)
Non-Hispanic Black	0 (0)	1 (0.8)
Hispanic	29 (21.5)	49 (36.8)
Asian or Pacific Islander	19 (14.1)	11 (8.3)
American Indian or Alaska Native	0 (0)	0 (0)
Other	19 (14.1)	17 (12.8)
Not stated	15 (11.1)	3 (2.3)
Insurance type, n (%)		
Private	104 (77.0)	83 (62.4)
Public	31 (23.0)	47 (35.3)
Both	0 (0)	2 (1.5)
No insurance	0 (0)	1 (0.8)
Primary language, n (%)		
English	117 (86.7)	112 (84.2)
Non-English	18 (13.3)	21 (15.8)

[a]TIDE: Timely Interventions for Diabetes Excellence.

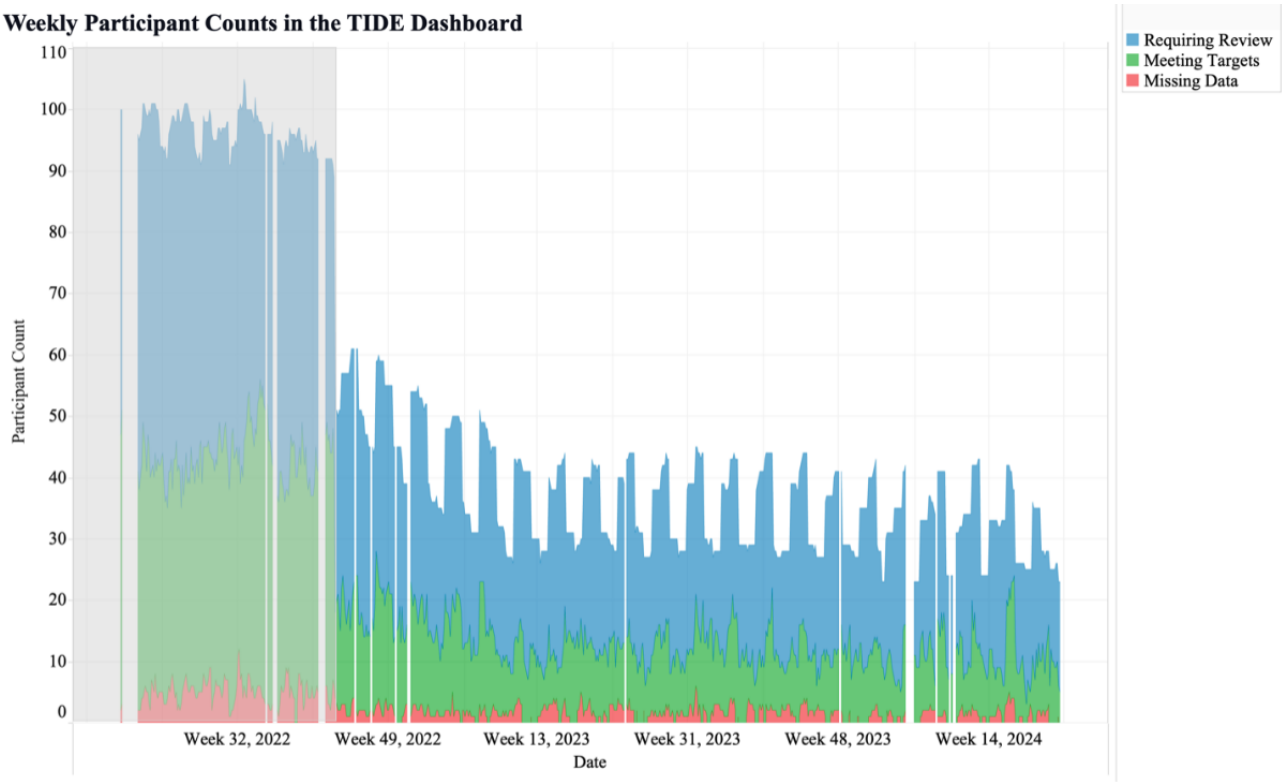
Unlike in traditional care models, there is no fixed number of participant counts. Participant counts change with glucose management, patients enrolling into the program, and patients transferring to other clinics, as well as the variable availability of the care providers, a format that has become more common in microrandomized trials [27]. Rigorous mathematical modeling to understand this is underway [28]. In visualizations for metrics 1-6, the gray area represents the duration of the 4T Pilot, and the white area represents the ongoing duration of the 4T Pilot long-term follow-up program.

Metrics and Visualizations for Clinical Workload

The first metric was the total number of participants shown in the TIDE dashboard. By monitoring the number of participants shown in TIDE, this metric allowed for the tracking of population statistics over time. The visualization of the metric selected by the CDCES team included the total number of participants in the program over time, segmented by those who did and did not require review (Figure 2). Based on this metric and visualization, the 4T Study team could investigate recruitment efforts or participant withdrawal rates if the metric was unexpectedly low or high. This metric was also used to understand the stringency of the criteria for participant review if the number of participants being reviewed was too high or too low.

XSL•FO
RenderX

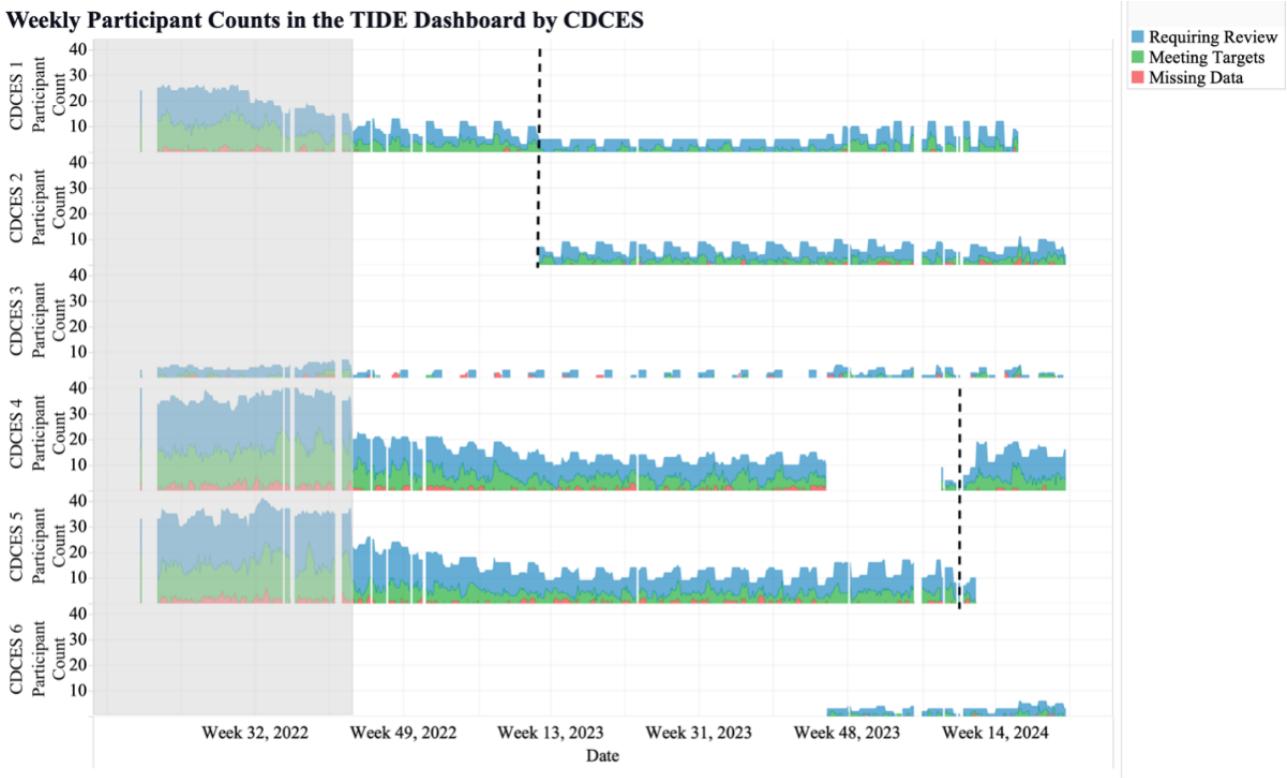
Figure 2. Weekly participant counts in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants requiring review (blue), meeting targets (green), and with more than 50% missing data (red). TIDE: Timely Interventions for Diabetes Excellence.



The second metric was the number of participants shown in TIDE, categorized by each CDCES in the program. CDCES names are numbered to protect anonymity. This metric revealed significant differences in the number of participants assigned to each CDCES, as well as stability in the number of participants reviewed by each CDCES over time (Figure 3). Differences in the number of patients assigned to each CDCES can be attributed to the level of their patients' complexity, their patients' primary languages, and each CDCES's time available to review participant data. This metric allowed for the monitoring of CDCES workload and identifying which CDCES team members may experience heavier participant workloads and require additional support, enabling workload balancing.

XSL•FO
RenderX

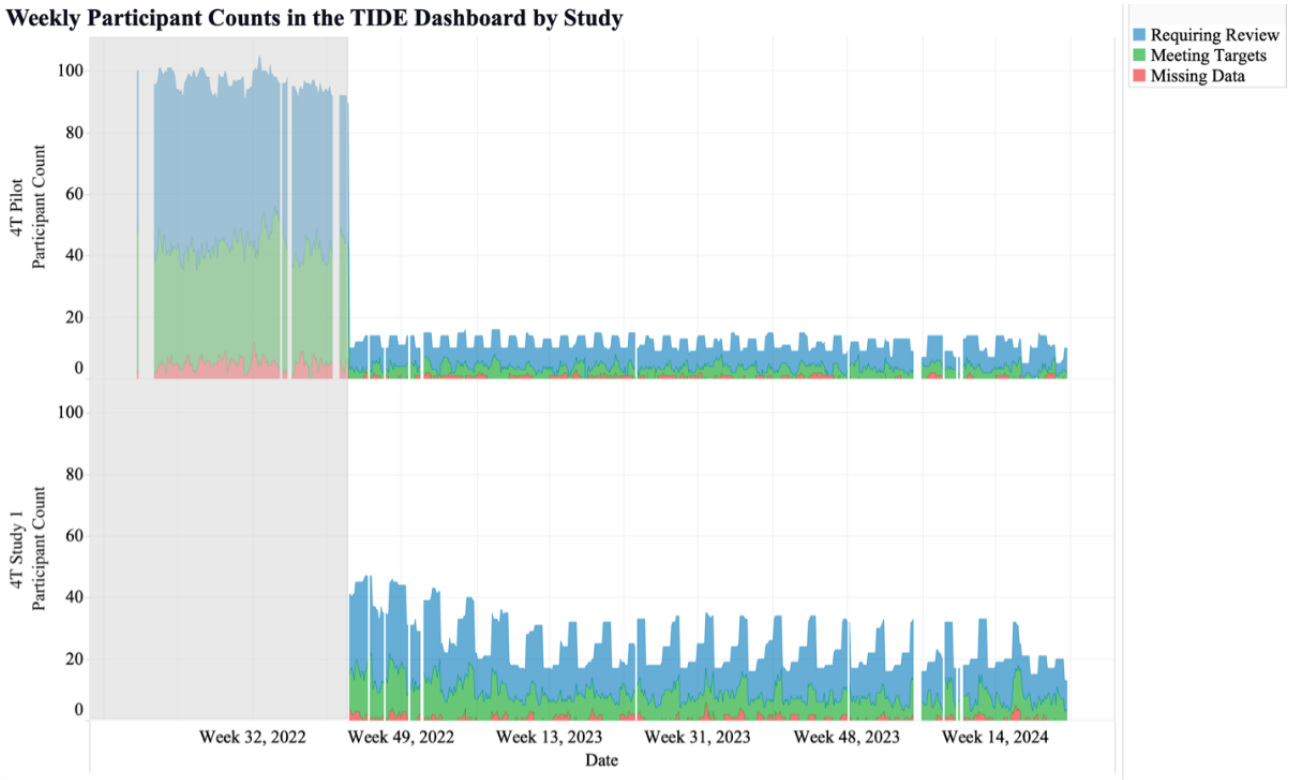
Figure 3. Weekly participant counts per Certified Diabetes Care and Education Specialist in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants requiring review (blue), meeting targets (green), and with more than 50% missing data (red). The dashed lines mark the transitions of participants between CDCES team members. CDCES: Certified Diabetes Care and Education Specialist; TIDE: Timely Interventions for Diabetes Excellence.



The third metric was the weekly number of participants shown in TIDE per study. This metric may help clinics with several studies view the number of participants in the RPM program for each study (Figure 4). Alternatively, this metric may be used for tracking the number of patients within various populations, such as the number of patients who are within 1 year of T1D diagnosis, at high risk for hypoglycemia, pregnant, non-English speaking, recently hospitalized, diagnosed with certain medical conditions or with certain family histories, using specific tools to manage their health, or pediatric versus adult patients.

XSL•FO
RenderX

Figure 4. Weekly participant counts per study in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants requiring review (blue), meeting targets (green), and with more than 50% missing data (red). TIDE: Timely Interventions for Diabetes Excellence.

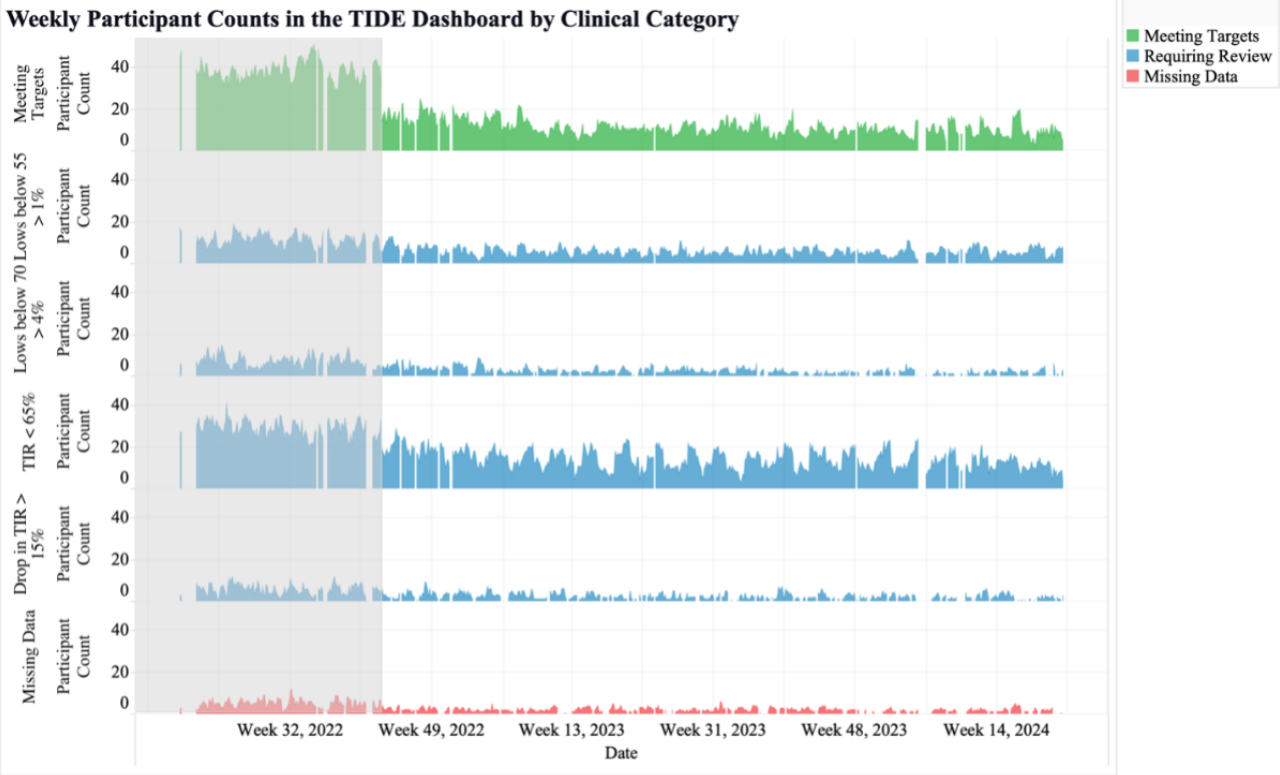


Metrics and Visualizations for Patient Outcomes

The fourth metric was the number of participants shown in TIDE per clinical category. This metric identified the level of variability in glucose management over time (Figure 5). This metric may help monitor participants' data and identify areas of improvement or concern, allowing CDCES team members to adjust care plans and interventions. For example, an increase in the number of participants missing glucose data may prompt CDCES team members to educate patients on how to improve connectivity with their CGM and how to upload glucose data.

XSL•FO
RenderX

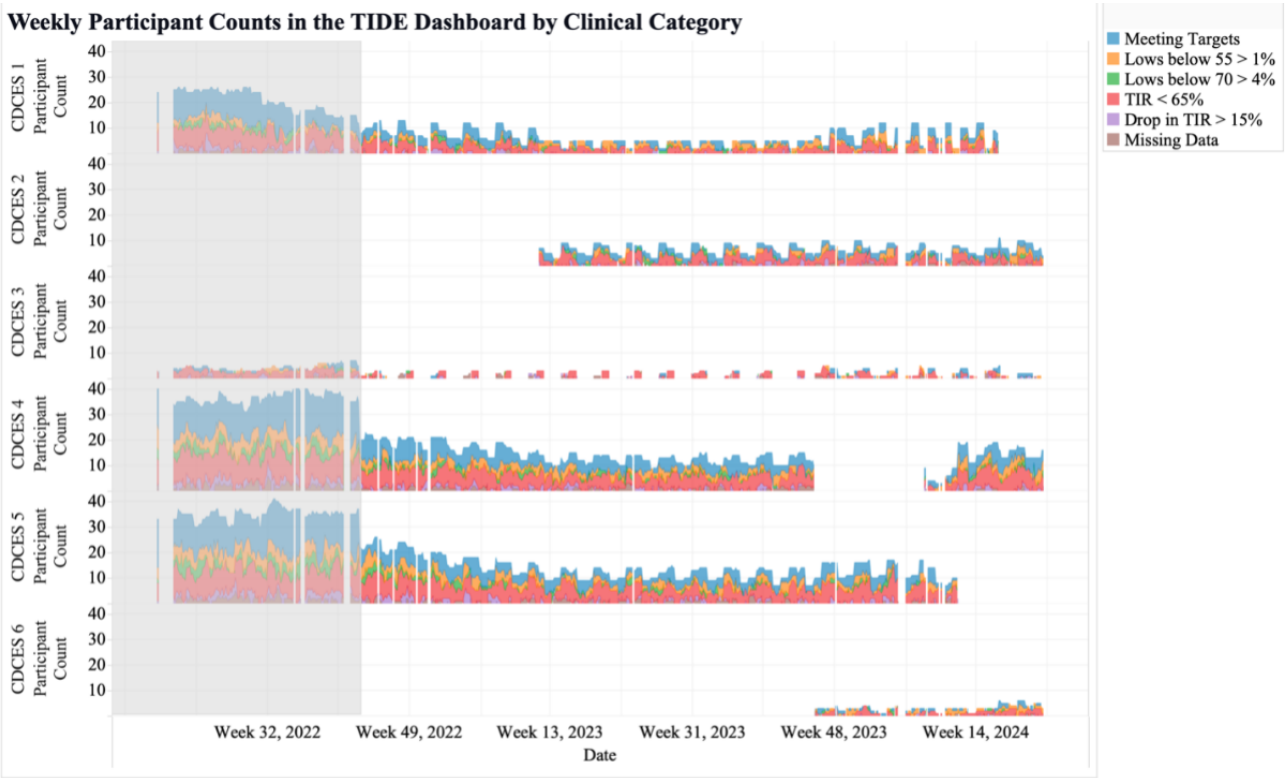
Figure 5. Weekly participant counts per clinical category in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants requiring review (blue), meeting targets (green), and with more than 50% missing data (red). TIDE: Timely Interventions for Diabetes Excellence; TIR: time in range.



The fifth metric was the number of participants shown in TIDE per clinical category and reviewing CDCES. This metric identified the similarities and differences in glucose management for the participants reviewed by each CDCES (Figure 6). In our clinic, the metric revealed that all CDCES team members had similar proportions of participants in each clinical category (Figure 6). This metric can track how well each CDCES's participants were meeting clinical consensus guidelines for CGM targets.

XSL•FO
RenderX

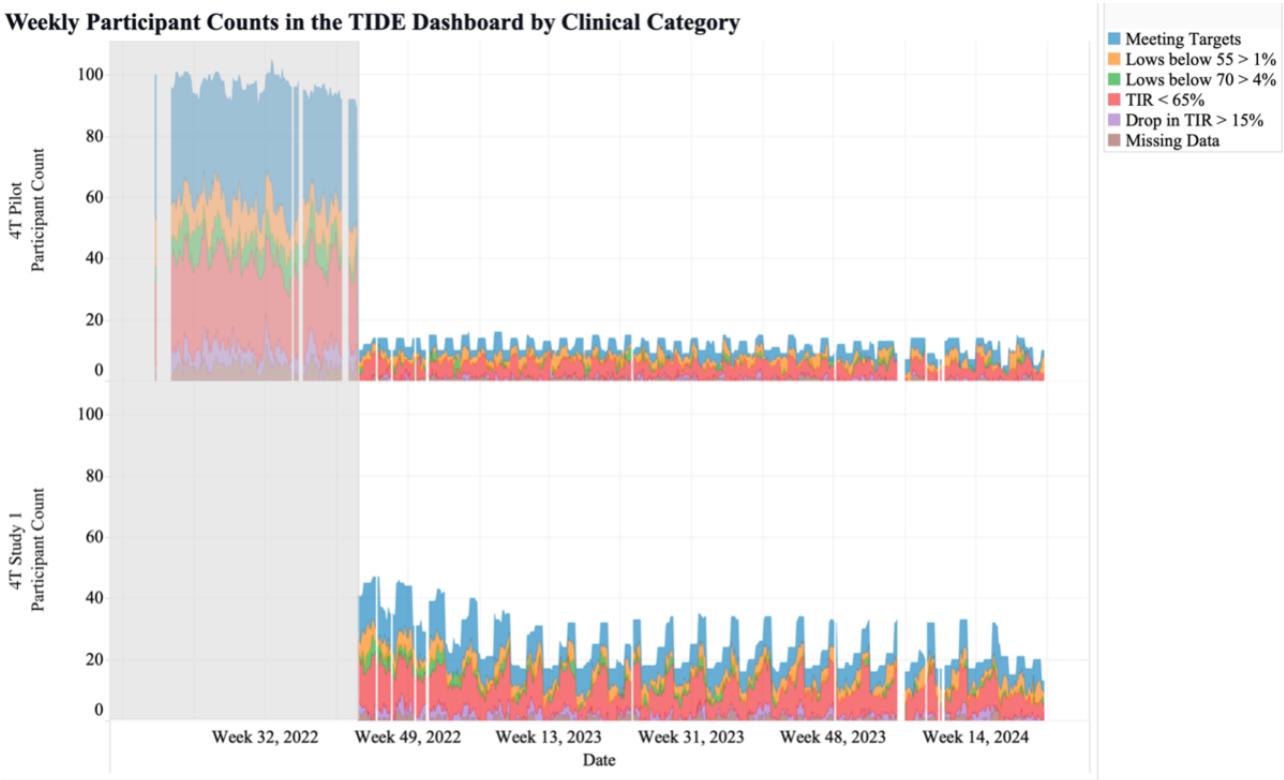
Figure 6. Weekly participant counts per Certified Diabetes Care and Education Specialist in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants meeting targets (blue), spending more than 1% of continuous glucose monitoring readings below 55 mg/dL (orange), spending more than 4% continuous glucose monitoring readings below 70 mg/dL (green), spending less than 65% of continuous glucose monitoring time in range (70 - 180 mg/dL; red), experiencing more than 15% point drop in time in range (purple), and missing more than 50% continuous glucose monitoring data (brown). CDCES: Certified Diabetes Care and Education Specialist; TIDE: Timely Interventions for Diabetes Excellence; TIR: time in range.



The sixth metric showed the weekly participant counts per study and clinical category. This metric may allow clinics to compare glucose management across studies or populations. This metric allowed the 4T team to identify variation in the numbers of participants in different studies and how often the groups were reviewed by CDCES team members (Figure 7). This provided insight into how well participants in different studies met clinical consensus glycemic targets. This metric facilitated easy comparison of health differences between different groups, study protocols (ie, Pilot vs Study 1), and clinical contact frequency (ie, weekly vs monthly contact).

XSL•FO
RenderX

Figure 7. Weekly participant counts per study in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants meeting targets (blue), spending more than 1% of continuous glucose monitoring readings below 55 mg/dL (orange), spending more than 4% continuous glucose monitoring readings below 70 mg/dL (green), spending less than 65% of continuous glucose monitoring time in range (70 - 180 mg/dL; red), experiencing more than 15% point drop in time in range (purple), and missing more than 50% continuous glucose monitoring data (brown). CGM: continuous glucose monitoring; TIDE: Timely Interventions for Diabetes Excellence; TIR: time in range.



Metric and Visualization for Timeliness of Clinical Reviews

The seventh metric was originally the number of days since each participant's most recent date of being displayed in TIDE. The patients shown were sorted from least to most recent display in TIDE and presented along with each participant's patient ID, the CDCES that reviewed their data, and the study they were enrolled in. After review of this metric at a 4T leadership meeting, several alternative metrics and visualizations were proposed. The chosen visualization was a table comprising participants who had not been shown for more than 20 days and ranked in the top 25% of patients with the longest time since being shown. These restrictions focused on the patients most likely to benefit from review and limited the number of patients to be reviewed by the CDCES team. In addition to showcasing the days since their last appearance in TIDE, the table included each participant's patient ID, the CDCES responsible for them, and the study they were enrolled in. If a participant had not been flagged in the RPM program in the last 20 days, this figure allowed the CDCES team to message or schedule a visit with participants, making sure no participants were inadvertently ignored. Similarly, this metric allowed clinical research coordinators to remove certain participants from studies that they were no longer enrolled in. For example, based on this metric, research coordinators checked if the participants who had not been shown in the TIDE dashboard for more than 50 days were still enrolled.

Feedback From the Care Team

The engineering team developed and implemented 7 operational metrics mapped to 3 KPI domains and embedded them in a dashboard used in regular leadership meetings. Administrative, clinical, and research feedback indicated the framework helped monitor clinical workload, patient outcomes, and timeliness of care.

Discussion

Principal Findings

We developed a quantitative framework for monitoring clinical workload, patient glucose management, and timeliness of care for a whole-population T1D care model in which an algorithm analyzes CGM data to help direct care providers to patients with CGM metrics not meeting targets. Through a multiyear process of data analysis, data visualization, and feedback from different team members, we identified metrics that informed oversight and decision-making for the algorithm-enabled care model. This framework may be helpful for other pediatric T1D clinics seeking to optimize their RPM-based care strategies around algorithms that help direct clinician efforts. This framework enables clinics to transition from fixed-interval models to data-driven, responsive care models. Although this work focuses on RPM in a pediatric T1D population, the concepts may translate to adult T1D populations and other chronic health conditions in which algorithms support and guide clinical care [29-31].

XSL•FO
RenderX

The current framework complements, rather than duplicates or substitutes, the way clinical trials report measures of workload, efficiency, and PROs and family-reported outcomes. It complements those measures by aggregating them at the clinic and population level for standardized review on an ongoing basis. For example, in the 4T Study, explicit measures of efficiency measured as per-patient time have been reported in studies describing years of work, while the current KPIs are reviewed on a regular basis [2,32].

The metrics developed yielded valuable insights into the RPM program and may provide insights into health disparities. First, our glucose management metrics—the fourth, fifth, and sixth metrics—reinforce the safety of our RPM platform and the larger 4T Study. As shown in Figure 5, flags related to time below range were minimal, while a significantly larger share resulted from patients meeting all glycemic targets. This aligns with our published findings, which demonstrate improved HbA_{1c} and TIR, minimal TBR, and positive PROs [9,17,22,33].

Second, despite efforts to evenly distribute CDCES workloads, the number of individuals meeting criteria for review varied during certain review periods, which may have resulted in less time dedicated to those individuals. This finding prompted the care team to consider redistributing subsequent enrollees. Third, the number of participants shown in TIDE per clinical category and reviewing CDCES (the fifth metric) may be used as a starting point to compare the performance of CDCES team members. Meaningful comparisons will require significant additional controls and considerations, such as patient complexity, engagement, and level of access to health care. Fourth, this tool has the potential to identify positive outliers (CDCES team members whose patients achieve clinical targets) or to monitor the results of pilot interventions deployed by a subset of the CDCESs.

Finally, this metric may be used as a tool to investigate health disparities and the impact of patient-provider identity congruence on the relationship between CDCESs and their participants. By stratifying metrics by study group and individual CDCES, we can identify disparities in care delivery and patient outcomes. Previous work in the 4T Pilot Study has already reported similar improvement in HbA_{1c} for Hispanic and non-Hispanic youth, as well as for publicly and privately insured youth, but it may be valuable to continually monitor these comparisons (and possibly patient primary language) as part of the KPI framework [34]. The algorithmic prioritization of patients may also limit clinician bias and subjective decision-making, as well as prioritize high-risk patients who may otherwise be overlooked. In the future, we may investigate the disparity in usage of diabetes technologies, incongruence in clinician and patient perceptions of diabetes technologies, as well as how interest in and feasibility of diabetes technology programs for patients with public insurance change over time.

The ability to track clinical categories can help inform decisions related to clinical efficiency and care delivery. Guided by insights from the fifth metric, CDCES team members could stage interventions to improve glucose management or CGM wear time, compare our clinic's TIR to other clinics, or observe how seasonal variations impact TIR.

The proposed framework is the first quantitative approach, of which we are aware, to monitor a CGM-based RPM program's outcomes, process metrics, and workloads. Developing metrics that drive decision-making is critical, particularly in an era where algorithms can significantly influence patient care and clinical workload. In our clinic, the review of these algorithm-enabled KPIs is becoming standard practice, integral to monitoring and adjusting a care model in which algorithms reduce clinical workload.

Recommendations to Implement KPIs and Metrics at a Clinic

KPI frameworks may be helpful for clinics that remotely access patient data, provide RPM-based care, and use algorithms to direct care delivery. Although our work focuses on RPM in a pediatric T1D population, these concepts may translate to an adult population, as well as other chronic health populations in which algorithms support clinical care. Using KPIs to monitor and adjust aspects of a care model may be particularly important for long-term interventions and large clinics or hospital systems [17].

To implement a KPI framework, the clinic or hospital might begin developing metrics by collaborating across multiple departments, potentially including internet technology specialists, data scientists, clinical research coordinators, clinicians, engineers, and quality improvement specialists. Clinics or hospitals may protect the time of or hire dedicated staff to implement the KPI framework and help with data analysis and visualization. We recommend that the internet technology team explore several platforms to host a dashboard for visualizing their metrics. Since large clinics may face scalability or data integration issues, it could be beneficial to partner with technology vendors or research collaborators. In a collaborative effort between Stanford's 4T team and Tidepool (a diabetes technology non-profit), a new clinic-agnostic, turnkey solution called Tidepool-TIDE was made available to any clinic in the United States [35]. This may help simplify the KPI framework implementation process for other clinics and hospitals.

Future Directions

Because of the success of our KPI framework, more metrics will be added in the future. We plan to add metrics related to CDCES time in TIDE, CDCES messaging activity, patient-clinician contact counts, PROs, and physical activity. These metrics have been researched as part of retrospective analyses, but it may be helpful to monitor them continuously as part of the KPI dashboard [2,7-9,18,22,24-26,36]. The above metrics have not yet been incorporated into the KPI framework because they are not integrated with the current workflow. We hope that the KPI framework supports the transition toward additional automation, such as using large language models to allow the platform to send some types of messages directly to the family [35,37]. The TIDE team continues to explore avenues for integration.

Limitations

We acknowledge that the 4T Pilot and Study 1 populations are not representative of all other pediatric diabetes clinics, but were

XSL•FO
RenderX

representative of the population diagnosed at our clinic [9,38]. This bias may limit the generalizability of our findings. Our study had high rates of early CGM adoption due to the design of the 4T intervention. Future implementation studies should include more representative samples outside of the context of broader research studies to validate the applicability of our framework across a wider range of clinical contexts. We also note that our study population is pediatric T1D, an autoimmune condition, in contrast to type 2 diabetes. T1D has the highest prevalence among non-Hispanic Whites, while type 2 diabetes has the highest prevalence among Native Americans and other non-White populations [39,40].

Although TIDE and the 4T Study have clearly improved patient outcomes and efficiency, we need to identify patients for whom repeated outreach is ineffective [2,8,9,17,21,22,25,26,32]. In future work, we will consider incorporating criteria to flag individuals who are repeatedly prioritized for review but are unreachable or do not seem to benefit from this prioritization. These patients may be switched from TIDE to a more traditional care model of quarterly reviews at clinic visits.

Conclusion

Data from CGMs, insulin pumps, and automated insulin delivery systems play a growing role in the management of T1D. To best use the data, clinics increasingly turn to trusted, transparent, and user-friendly algorithms to translate patient data into insights that inform patient care. As the role of algorithms grows, the proposed framework may offer clinics a quantitative approach to monitor and potentially adjust the care delivery model.

Acknowledgments

The authors would like to acknowledge all the youth who participated in the 4T Study. We would like to thank the other members of the research team, including research coordinators, clinical staff, students in the Systems Utilization Research for Stanford Medicine (SURF) group, the Quantitative Sciences Unit, the T1D Working Group in Statistics, and Informatics at Stanford Medicine Children's Health.

Funding

This work was supported by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; R18DK122422), the Helmsley Charitable Trust (G-2002-04251- 2), the ISPAD-JDRF Research Fellowship (SDRC; 1P30DK 11607401), LPCH Auxiliaries, the National Science Foundation (2205084), Stanford HAI, and Stanford REDCap Platform (UL1 TR003142).

Authors' Contributions

Conceptualization: JOF (equal), PD (equal), DMM (supporting), PP (supporting), DS (equal)
Software: JOF (supporting), PD (supporting)
Review and editing: DPZ (supporting), DMM (supporting), PP (supporting), RJ (supporting), FKB (supporting), AA (supporting), DS (equal)
Methodology: DPZ (supporting), DMM (supporting), PP (supporting), RJ (supporting), FKB (supporting), AA (supporting), DS (equal)
Writing – original draft: JK (lead)
Writing – review & editing: JK (lead), DS (supporting)
All roles: JK (lead)

Conflicts of Interest

DMM, DS, PP, and RJ have received support from Stanford MCHRI, Stanford HAI, and the NSF. DS, RJ, DMM, DPZ, and AA have received funding from the Helmsley Charitable Trust. DPZ has received honoraria for speaking engagements from Ascensia Diabetes, Insulet Canada, Medtronic Diabetes, and Dexcom; research support from the ISPAD-JDRF Research Fellowship; and serves on the advisory board of DexCom Inc. AA has received research support from the NIH, MCHRI, and grants K12DK122550 (Stanford University) and K23DK131342 from the NIDDK during the conduct of the study. DMM has had research support from the NIDDK, NIH, Breakthrough T1D, and grant number P30DK116074; his institution has had research support from DexCom Inc., Medtronic, Insulet, Bigfoot Biomedical, Tandem, and Roche; and has consulted for Abbott, the Helmsley Charitable Trust, Lifescan, Sanofi, Medtronic, Provention Bio, Kriya, Novo Nordisk, Eli Lilly, Insulet, Biospex, and Bayer. JF has received support from an NSF grant. DS is an adviser to Carta Health. JK interned at Dexcom. All other authors declare that they have no competing interests.

Multimedia Appendix 1
Additional information on interviews, metric visualizations, and insights.
[DOCX File, 18 KB - diabetes_v11i1e72676_app1.docx]

References

XSL•FO
RenderX

1. Su D, Michaud TL, Estabrooks P, et al. Diabetes management through remote patient monitoring: the importance of patient activation and engagement with the technology. Telemed J E Health 2019 Oct;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
2. Ferstad JO, Vallon JJ, Jun D, et al. Population-level management of type 1 diabetes via continuous glucose monitoring and algorithm-enabled patient prioritization: precision health meets population health. Pediatr Diabetes 2021 Nov;22(7):982-991. [doi: 10.1111/pedi.13256] [Medline: 34374183]
3. American Diabetes Association Professional Practice Committee. 7. Diabetes Technology: Standards of Medical Care in Diabetes-2022. Diabetes Care 2022 Jan 1;45(Suppl 1):S97-S112. [doi: 10.2337/dc22-S007] [Medline: 34964871]
4. Andrès E, Meyer L, Zulfiqar AA, et al. Telemonitoring in diabetes: evolution of concepts and technologies, with a focus on results of the more recent studies. J Med Life 2019;12(3):203-214. [doi: 10.25122/jml-2019-0006] [Medline: 31666818]
5. Crossen SS, Romero CC, Lewis C, Glaser NS. Remote glucose monitoring is feasible for patients and providers using a commercially available population health platform. Front Endocrinol (Lausanne) 2023;14:1063290. [doi: 10.3389/fendo.2023.1063290] [Medline: 36817610]
6. Crossen S, Romero C, Reggiardo A, Michel J, Glaser N. Feasibility and impact of remote glucose monitoring among patients with newly diagnosed type 1 diabetes: single-center pilot study. JMIR Diabetes 2022 Jan 17;7(1):e33639. [doi: 10.2196/33639] [Medline: 35037887]
7. Prahalad P, Ding VY, Zaharieva DP, et al. Teamwork, Targets, Technology, and Tight control in newly diagnosed type 1 diabetes: the Pilot 4T Study. J Clin Endocrinol Metab 2022 Mar 24;107(4):998-1008. [doi: 10.1210/clinem/dgab859] [Medline: 34850024]
8. Prahalad P, Tanenbaum M, Hood K, Maahs DM. Diabetes technology: improving care, improving patient-reported outcomes and preventing complications in young people with type 1 diabetes. Diabet Med 2018 Apr;35(4):419-429. [doi: 10.1111/dme.13588] [Medline: 29356074]
9. Prahalad P, Scheinker D, Desai M, et al. Equitable implementation of a precision digital health program for glucose management in individuals with newly diagnosed type 1 diabetes. Nat Med 2024 Jul;30(7):2067-2075. [doi: 10.1038/s41591-024-02975-y] [Medline: 38702523]
10. Kirkland EB, Marsden J, Zhang J, et al. Remote patient monitoring sustains reductions of hemoglobin A1c in underserved patients to 12 months. Prim Care Diabetes 2021 Jun;15(3):459-463. [doi: 10.1016/j.pcd.2021.01.005] [Medline: 33509728]
11. Lacy ME, Lee KE, Atac O, et al. Patterns and trends in continuous glucose monitoring utilization among commercially insured individuals with type 1 diabetes: 2010-2013 to 2016-2019. Clin Diabetes 2024;42(3):388-397. [doi: 10.2337/cd23-0051] [Medline: 39015169]
12. Scheinker D, Gu A, Grossman J, et al. Algorithm-enabled, personalized glucose management for type 1 diabetes at the population scale: prospective evaluation in clinical practice. JMIR Diabetes 2022 Jun 6;7(2):e27284. [doi: 10.2196/27284] [Medline: 35666570]
13. Thapa I, Laforcade PA, Bishop FK, et al. Regression discontinuity in time: evaluating the impact of evolving digital health interventions. Int J Med Inform 2025 Dec;204:106050. [doi: 10.1016/j.ijmedinf.2025.106050] [Medline: 40729778]
14. Gartner JB, Lemaire C. Dimensions of performance and related key performance indicators addressed in healthcare organisations: a literature review. Int J Health Plann Manage 2022 Jul;37(4):1941-1952 [FREE Full text] [doi: 10.1002/hpm.3452] [Medline: 35288968]
15. CBDCE brochure. : Certification Board for Diabetes Care and Education; 2021 URL: https://www.cbdce.org/documents/20123/0/Final+CBDCE_Brochure_CDCES_2021_Rev.0321.pdf/526c9c85-5b04-c3b5-eeb9-4b500d4ad55b?t=1619727057105 [accessed 2025-09-05]
16. Leverenz JC, Horton L, Conrad B, et al. Insulin dose adjustment policy for certified diabetes care and education specialists: safe and effective. J Clin Transl Endocrinol 2025 Dec;42:100423. [doi: 10.1016/j.jcte.2025.100423] [Medline: 41283065]
17. Prahalad P, Ding VY, Zaharieva DP, et al. Sustained HbA1c improvements over 36 months in youth in the Teamwork, Targets, Technology, and Tight Control (4T) Study. J Clin Endocrinol Metab 2026 Jan 21;111(2):418-426. [doi: 10.1210/clinem/dgaf397]
18. Zaharieva DP, Bishop FK, Maahs DM, 4T study research team. Advancements and future directions in the teamwork, targets, technology, and tight control-the 4T study: improving clinical outcomes in newly diagnosed pediatric type 1 diabetes. Curr Opin Pediatr 2022 Aug 1;34(4):423-429. [doi: 10.1097/MOP.0000000000001140] [Medline: 35836400]
19. Chang A, Gao MZ, Ferstad JO, et al. A quantitative model to ensure capacity sufficient for timely access to care in a remote patient monitoring program. Endocrinol Diabetes Metab 2023 Sep;6(5):e435. [doi: 10.1002/edm2.435] [Medline: 37345227]
20. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
21. Tanenbaum ML, Addala A, Hanes S, et al. “It changed everything we do”: a mixed methods study of youth and parent experiences with a pilot exercise education intervention following new diagnosis of type 1 diabetes. J Diabetes Complicat 2024 Jan;38(1):108651. [doi: 10.1016/j.jdiacomp.2023.108651]

XSL•FO
RenderX

22. Addala A, Ritter V, Schneider-Utaka AK, et al. Psychosocial outcomes in a diverse sample of youth and their families who initiated continuous glucose monitoring within the first year of type 1 diabetes diagnosis. Diabetes Obes Metab 2025 Feb;27(2):933-943. [doi: 10.1111/dom.16093] [Medline: 39604317]
23. Wright EE, Morgan K, Fu DK, Wilkins N, Guffey WJ. Time in range: how to measure it, how to report it, and its practical application in clinical decision-making. Clin Diabetes 2020 Dec;38(5):439-448. [doi: 10.2337/cd20-0042] [Medline: 33384469]
24. Tanenbaum ML, Zaharieva DP, Addala A, et al. “Much more convenient, just as effective”: experiences of starting continuous glucose monitoring remotely following type 1 diabetes diagnosis. Diabet Med 2022 Nov;39(11):e14923. [doi: 10.1111/dme.14923] [Medline: 35899591]
25. Tanenbaum ML, Zaharieva DP, Addala A, et al. “I was ready for it at the beginning”: parent experiences with early introduction of continuous glucose monitoring following their child’s type 1 diabetes diagnosis. Diabet Med 2021 Aug;38(8):e14567. [doi: 10.1111/dme.14567] [Medline: 33772862]
26. Bishop FK, Chmielewski A, Leverenz J, et al. Building a diabetes educator program for remote patient monitoring billing. J Diabetes Sci Technol 2024 Dec 29:19322968241308920. [doi: 10.1177/19322968241308920] [Medline: 40438903]
27. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
28. Ferstad JO, Prahalad P, Maahs DM, et al. Smart Start - designing powerful clinical trials using pilot study data. NEJM Evid 2024 Feb;3(2):EVIDoa2300164. [doi: 10.1056/EVIDoa2300164] [Medline: 38320487]
29. Singh N, Lawrence K, Andreadis K, Twan C, Mann DM. Developing and scaling remote patient monitoring capacity in ambulatory practice. NEJM Catalyst 2024 May 15;5(6). [doi: 10.1056/CAT.23.0417]
30. Pritchett JC, Paludo JC, Borah BJ, Haddad TC. Remote patient monitoring to transform management of febrile neutropenia in cancer patients. NEJM Catalyst 2024 Mar 20;5(4). [doi: 10.1056/CAT.23.0365]
31. Graseck A, Josephs M, Dokras A, Srinivas SK. THEA: prenatal education and remote blood pressure monitoring. NEJM Catalyst 2024 Jan 17;5(2). [doi: 10.1056/CAT.23.0343]
32. Scheinker D, Prahalad P, Johari R, Maahs DM, Majzun R. A new technology-enabled care model for pediatric type 1 diabetes. NEJM Catalyst 2022 Apr 20;3(5). [doi: 10.1056/CAT.21.0438]
33. Tanenbaum ML, Pang E, Tam R, et al. “We’re taught green is good”: perspectives on time in range and time in tight range from youth with type 1 diabetes, and parents of youth with type 1 diabetes. Diabet Med 2024 Dec;41(12):e15423. [doi: 10.1111/dme.15423] [Medline: 39118381]
34. Addala A, Ding V, Zaharieva DP, et al. Disparities in hemoglobin A1c levels in the first year after diagnosis among youths with type 1 diabetes offered continuous glucose monitoring. JAMA Netw Open 2023 Apr 3;6(4):e238881. [doi: 10.1001/jamanetworkopen.2023.8881] [Medline: 37074715]
35. Kim GYK, Rostosky R, Bishop FK, et al. The adaptation of a single institution diabetes care platform into a nationally available turnkey solution. NPJ Digit Med 2024 Nov 6;7(1):311. [doi: 10.1038/s41746-024-01319-x] [Medline: 39506045]
36. Zaharieva DP, Senanayake R, Brown C, et al. Adding glycemic and physical activity metrics to a multimodal algorithm-enabled decision-support tool for type 1 diabetes care: keys to implementation and opportunities. Front Endocrinol (Lausanne) 2022;13:1096325. [doi: 10.3389/fendo.2022.1096325] [Medline: 36714600]
37. Kurtzig J, Agrawal A, Tascioglu M, Goodyear L. A. ATTD 2025 Amsterdam & Online—19–22 March and ATTD-ASIA 2024 Singapore—18–20 November. Diabetes Technol Ther 2025 Mar 1;27(S2):e1-e441. [doi: 10.1089/dia.2024.78502.abstracts]
38. Nally LM, Sherr JL. Making diabetes technology accessible to all. Nat Med 2024 Jul;30(7):1832-1833. [doi: 10.1038/s41591-024-03082-8] [Medline: 38918631]
39. Yedjou CG, Sims JN, Njiki S, et al. Health and racial disparities in diabetes mellitus prevalence, management, policies, and outcomes in the United States. J Community Med Public Health 2024;8(3):460. [doi: 10.29011/2577-2228.100460] [Medline: 39697180]
40. Fang M, Wang D, Selvin E. Prevalence of type 1 diabetes among US children and adults by age, sex, race, and ethnicity. JAMA 2024 Apr 23;331(16):1411-1413. [doi: 10.1001/jama.2024.2103] [Medline: 38573653]

Abbreviations

4T: Teamwork, Targets, Technology, and Tight Control
CDCES: Certified Diabetes Care and Education Specialist
CGM: continuous glucose monitor
EHR: electronic health record
HbA_{1c}: hemoglobin A_{1c}
KPI: key performance indicator
PRO: patient-reported outcome
REDCap: Research Electronic Data Capture
RPM: remote patient monitoring

XSL•FO
RenderX

T1D: type 1 diabetes
TBR: time below range
TIDE: Timely Interventions for Diabetes Excellence
TIR: time in range

Edited by A Stone; submitted 15.Feb.2025; peer-reviewed by J Ackerman, R Shulman; revised version received 02.Feb.2026; accepted 02.Feb.2026; published 25.Mar.2026.

Please cite as:
Kurtzig J, Addala A, Bishop FK, Dupenloup P, Ferstad JO, Johari R, Maahs DM, Prahalad P, Zaharieva DP, Scheinker D
A Quantitative Framework for Evaluating the Performance of Algorithm-Directed Whole-Population Remote Patient Monitoring: Tutorial for Type 1 Diabetes Care
JMIR Diabetes 2026;11:e72676
URL: https://diabetes.jmir.org/2026/1/e72676
doi:10.2196/72676

© Jamie Kurtzig, Ananta Addala, Franziska K Bishop, Paul Dupenloup, Johannes O Ferstad, Ramesh Johari, David M Maahs, Priya Prahalad, Dessi P Zaharieva, David Scheinker. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 25.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) and Hospital-Based Approaches to Closed-Loop Therapy in Type 1 Diabetes: Cost-Effectiveness Analysis

Mercia Napame[1], MHE; Sylvie Picard[2], MD, PhD; Tony Foglia[3], PhD; Anne Rubenstrunk[1], PhD; Florence Baudoux[1], MD; Francoise Giroud[4], MD; Sandrine Lablanche[5], MD, PhD; Sophie Borot[6], MD, PhD

[1]Santelys, Loos, France
[2]CIRDIA, Point Medical, Rond-Point de la Nation, Dijon, France
[3]URPS-ML-BFC, Burgundy-Europe University, Dijon, France
[4]CDOM21, CIRDIA, Dijon, France
[5]Centre Hospitalier Universitaire de Grenoble, Grenoble, France
[6]Centre Hospitalier Universitaire de Besançon, Besançon, France

Corresponding Author:
Sylvie Picard, MD, PhD
CIRDIA, Point Medical, Rond-Point de la Nation, Dijon, France

Abstract

Background: Closed-loop insulin delivery is the new standard of care for patients with type 1 diabetes (T1D). However, in France, its implementation remains predominantly hospital based. Expanding access to this treatment through alternative care models looks essential.

Objective: This study (cost-effectiveness analysis) compares 2 care models for people with T1D implementing a closed-loop system in France: outpatient care in the Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) and inpatient care.

Methods: We conducted a cost-effectiveness analysis using retrospective observational data from individuals with T1D aged 16 years and older from the implementation of the closed loop to a 12-month follow-up either in the CIRDIA (CIRDIA group) or in a hospital center setting (hospital center [HC] group). The cost analyses were based on patient records and public databases: the French Medical Information Systems Program and the French General Nomenclature of Professional Acts. Closed-loop efficacy was assessed using a time in range (TIR) of 70 to 180 mg/dL, and closed-loop safety was assessed using the glycemia risk index (GRI), a single indicator that represents the risk of hypoglycemia or hyperglycemia and ranges from 0 (minimal risk) to 100 (maximal risk).

Results: A total of 201 patients were included: 128 in the CIRDIA group and 73 in the HC group. The mean (SD) age was 43 (14) years and 46 (15) years, respectively. Mean (SD) baseline TIR was 52.9% (16%) in the CIRDIA group versus 65.9% (15.1%) in the HC group ($P<.001$), whereas mean (SD) baseline GRI was 56.4 (21) in the CIRDIA group versus 37.8 (19.8) in the HC group ($P<.001$). After 12 months, both groups achieved similar efficacy and safety outcomes with a mean (SD) TIR at 72.7% (11.6%) in the CIRDIA group versus 71.9% (10.5%) in the HC group, and a mean GRI at 30.1 (14.1) versus 30.3 (13), respectively. There were no significant between-group differences ($P=.60$ for TIR; $P=.91$ for GRI). However, the CIRDIA was associated with significantly lower management costs with a mean cost of €8373.12 (SD €427.30; €1=US $1.10 at the time of the study) per patient in the CIRDIA group versus €8814.32 (SD €192) per patient in the HC group ($P<.001$). The estimated saving was €626 per percentage point of increase in TIR and €2011 per point of reduction in GRI, indicating that the HC closed-loop initiation was dominated by the CIRDIA. The CIRDIA was less costly than HC in 8600 (86%) out of 10,000 simulations in a probabilistic sensitivity analysis.

Conclusions: These findings suggest the potential of the CIRDIA to represent a viable alternative organizational model for closed-loop initiation in France, achieving comparable effectiveness at lower cost in our population. Further research with longer follow-up is warranted. From a policy perspective, the resources saved could be at least partly reallocated to support out-of-hospital closed-loop initiation centers.

(JMIR Diabetes 2026;11:e86690) doi:10.2196/86690

KEYWORDS

closed-loop; cost-effectiveness analysis; follow-up studies; hospital; treatment outcome; type 1 diabetes

XSL•FO
RenderX

Introduction

Background

Diabetes is a chronic disease characterized by persistent hyperglycemia, resulting from either a relative or absolute deficiency in insulin secretion or an impairment in its action. It represents a major public health challenge because of its increasing prevalence, its impact on patients' quality of life, and the substantial economic burden on health care systems [1]. As of 2024, 588.8 million adults (aged 20 - 79) worldwide were living with diabetes, a number projected to increase to approximately 853 million by 2050 [2]. In France, more than 4.5 million people are living with diabetes [3]. Among the different forms of diabetes, type 1 diabetes (T1D) is an autoimmune disease that is often diagnosed in children, adolescents, or young adults. Overall, about 7.4 million people are living worldwide with T1D, and in France, T1D accounts for approximately 320,000 individuals [3]. The management of T1D requires lifelong insulin therapy, frequent or nowadays continuous glucose monitoring (CGM), and structured patient therapeutic education [4]. Over the past decades, technological advances have progressively transformed diabetes care from multiple daily injections to external insulin pumps and subsequently to CGM, enabling the real-time tracking of glycemia [5]. These innovations have paved the way for the development of closed-loop (CL) systems, which integrate a glucose sensor, an insulin pump, and an adaptive control algorithm [5].

Prior Work

While numerous studies have established the clinical benefits of CL systems on glycemic outcomes, evidence on the models of care for their initiation and follow-up remains limited [6-10]. The recent reimbursement of CL systems in France, and the relative novelty of studying organizational rather than purely clinical outcomes, may explain this evidence gap [11].

In France, approximately 2 years after the first reimbursement, only about 15,000 eligible patients had received CL systems—roughly a 5% coverage—despite the benefits for glycemic control [12]. This low rate is partly attributable to the centralization of CL initiation in hospital-based clinics, where waiting times are often long [13].

The Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) was developed in 2023 mainly to improve access to CL among persons with T1D. The CIRDIA is a multisite CL initiation center regrouping highly trained diabetologists, mostly in private practice. The CIRDIA—like hospital-based CL initiation centers—is based on the guidelines of the French-Speaking Diabetes Society (SFD) [4]. However, as this is a new concept of care in France, its cost-effectiveness had to be evaluated and compared to usual hospital-based care.

Study Objectives

Evidence on the cost-effectiveness of alternative organizational models of CL initiation, such as out-of-hospital–based pathways, remains scarce. This raises the question of whether initiating CL systems in out-of-hospital settings, such as the CIRDIA, could represent a cost-effective alternative to hospital center (HC)–based initiation.

This study aimed to estimate the 1-year cost-effectiveness of CIRDIA-based CL initiation compared to HC-based initiation among patients with T1D in France from a French National Health Insurance perspective. We hypothesized that out-of-hospital–based initiation could achieve comparable effectiveness and safety while reducing costs. Evaluating this organizational model could determine whether or not the CIRDIA represents a viable alternative for the French health care system and provide the data that may be transferable to other health care systems worldwide.

Methods

Study Design

This is a cost-effectiveness analysis based on retrospective observational data collected between 2023 and 2024 with a 12-month follow-up as part of the routine monitoring of patients with T1D initiating CL in France. We compared 2 modes of health care delivery: the CIRDIA setting and the HC setting. The cost-effectiveness analysis compared the net monetary costs of health care intervention with a measure of its clinical effectiveness.

Accordingly, the evaluation was conducted from the perspective of the French National Health Insurance (Assurance Maladie), considering all costs covered by the payer, with a 1-year time horizon. No modeling was conducted, as all analyses relied on real-world data extracted from patient records (follow-up consultations) and public databases: the Agency for Information on Hospital Care and the French Health Insurance [14].

Recruitment

The study included persons living with T1D, 16 years of age or older, starting for the first time a CL system. Patients with missing continuous glucose monitoring data were excluded. Participants were allocated to 1 of the 2 groups based on their care pathway: those managed directly by the CIRDIA center (CIRDIA group) and those initiated and followed by the hospital center outpatient clinic (HC group). The 2 models of care were mutually exclusive and could not be used simultaneously.

Participants from the CIRDIA group were consecutive patients who started CL between May 2, 2023, and March 30, 2024, and had at least a 12-month follow-up. Devices (insulin pump, infusion sets, insulin reservoirs, and glucose sensors) were provided by different home health care providers, as it is the rule in France. Registered nurses specialized in diabetes care and working for home health care providers are usually responsible for the technical education of the patient and connectivity issues. Participants in the HC group had CL initiated in 2023 or 2024 in 1 of the 5 HCs located in the north of France ("Haut-de-France" region) and were the patients for whom devices and technical education were provided by Santelys, a nonprofit organization acting as a home health care provider.

XSL•FO
RenderX

Ethical Considerations

This study used retrospective observational data collected as part of the routine monitoring of persons with T1D managed on CL therapy. No additional intervention occurred beyond usual care. All data were fully anonymized before analysis in accordance with the General Data Protection Regulation. No patient could be identified directly or indirectly [15]. In line with current regulations regarding research not involving human persons, no specific ethics committee approval was required [16].

All participants had received oral and written information at the time of CL initiation about the potential use of their anonymized clinical data for research purposes. Written consent or non-opposition was obtained in accordance with French data protection and ethical regulations. This study complied with the principles of the Declaration of Helsinki and relevant national guidelines regulating the secondary use of health data.

Interventions

The CIRDIA is a new concept in France of a multisite health care model that performs CL initiation most often during a long (about 1 h) office visit or occasionally during a day hospitalization (DH) outside of university hospitals. Its activity complies with the position statement issued by the SFD and the French National Health Authority (HAS) [17]. The main objective of the CIRDIA is to expand access to care for people living with T1D while reducing the burden on HC. Furthermore, initiating CL systems in the out-of-hospital sector is considered a strategic lever to support the sustainability of out-of-hospital diabetes care. Nevertheless, since CL initiation is predominantly performed in hospital settings, hospital-based care is considered the reference strategy. The out-of-hospital sector initiation remains underdeveloped and must demonstrate its effectiveness.

In the CIRDIA arm, CL initiation was usually followed by 3 teleconsultations and 3 consultations over 1 year. For some patients (those initiated after January 1, 2024), an additional 3-month telemonitoring period could be implemented. In the HC arm, CL initiation was carried out during DH, followed by 3 teleconsultations and 3 follow-up visits, coupled with 3 months of telemonitoring for patients initiated after January 1, 2024. In both settings, CGM data were available for the diabetologist (or the diabetes care team) to optimize patient adherence to the device [18].

Efficacy and Safety Inputs

Because CL initiation and the 1-year time horizon did not affect mortality or lifespan, we selected an alternative measure for effectiveness. However, due to incomplete data on comorbidities and complications in 1 of the 2 study arms (HC), adverse events could not be included in the analysis. Instead, effectiveness was assessed by improvement in the time in range (TIR) 70 - 180 mg/dL, while safety was assessed through a reduction in the glycemia risk index (GRI). The GRI is a composite metric that reflects both hypoglycemia and hyperglycemia risks by integrating the time spent below range (<54 mg/dL and 54 - 69 mg/dL) and the time spent above range (181 - 250 mg/dL and >250 mg/dL). Notably, although hemoglobin A_{1c} is frequently used as an efficacy outcome in similar studies, it is no longer systematically measured during routine consultations [19].

Cost Inputs

We conducted the economic evaluation from a health care payer perspective, including all direct medical and nonmedical expenses reimbursed by the French National Health Insurance, expressed in euros for the year 2024. Costs were estimated using a bottom-up micro-costing approach, which is considered the gold standard in health technology cost assessment according to HAS recommendations. Because T1D belongs to the list of fully covered diseases by the French National Health Insurance, no out-of-pocket expense was considered. Moreover, because the time horizon was limited to 1 year, no discount rate was applied. Cost components were identified and calculated in line with the HAS and SFD recommendations [20,21].

Outpatient procedures and consultations were valued according to the prices from the General Classification of Professional Acts and the Common Classification of Medical Acts. Biological analyses were valued according to the Common Nomenclature of Medical Biology Acts. In addition, CL-related costs were valued in accordance with the List of Products and Services of the French National Health Insurance. The cost of DH was calculated using the Homogeneous Group of Patients with the principal diagnosis code Z451 ("Adjustment and maintenance of an infusion pump"), associated with the Hospital Stay Tariff 1794, based on prices provided by the Agency for Information on Hospital Care [22-26].

Incremental Cost-Effectiveness Ratio

The results of a cost-effectiveness analysis were expressed in terms of incremental cost-effectiveness ratios (ICERs) and were calculated as the ratio of incremental costs to incremental health outcomes between the 2 groups. Specifically, ICERs were expressed as the additional cost per percentage point of increase in TIR and per unit of reduction in the GRI. In line with International Society for Pharmacoeconomics and Outcomes Research recommendations, negative ICERs were interpreted as situations of dominance or dominated strategies rather than reported as such. A strategy was considered to be dominated if it was more costly and less effective or more costly and equally effective. We designed, conducted, and reported this evaluation in accordance with the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) guidelines [27].

Sensitivity Analysis

As this study was based on real-world observational data rather than modeled parameters, some uncertainty may still arise from the data, potentially leading to biased estimates. According to the International Society for Pharmacoeconomics and Outcomes Research [28], deterministic sensitivity analysis was not applicable in this context. Instead, robustness was explored through subgroup analyses and through a probabilistic sensitivity analysis (PSA) to test whether the conclusions of the base-case analysis held under parameter uncertainty. A PSA was performed using 10,000 Monte Carlo simulations in which all parameters were varied simultaneously. Parameter values were sampled from predefined probability distributions: truncated

XSL•FO
RenderX

normal for efficacy and safety outcomes (bounded between 0 and 100) and gamma for costs [28].

Statistical Analysis

Data were collected using Excel (2016, Microsoft Inc.), and statistical analyses were performed with the R software version 4.4.2 (2024). Means and SDs were calculated for quantitative variables. To verify comparability between the groups, we conducted a Shapiro-Wilk test to check the normality of our variables. For normally distributed variables, we used a 2-tailed Student *t* test, and for non-normally distributed variables, the Wilcoxon signed-rank test was used. The threshold for statistical significance was set at $P<.05$.

Results

Overview

Overall, 201 patients aged 16 to 80 years were included in this study, including 128 CL initiations by the CIRDIA and 73 by HC. Baseline characteristics of the 2 groups are shown in Table 1. The mean age of patients initiated at CL by the CIRDIA was 43 (SD 14) years and 46 (SD 15) years for the HC arm. The gender distribution was 52% (n=66 and n=38) women and 48% (n= 62 and n=35) men in both arms, and the average BMI was 27.5 (SD 4.9) and 27.2 (SD 5.2) kg/m^2, respectively. In the CIRDIA arm, only 17 (13%) CL initiations were performed during DH, while the remaining initiations were conducted during 1-hour office visits.

Table . Baseline characteristics of the patients included in the study by group (CIRDIA[a] vs HC[b]).

Parameters	CIRDIA (n=128)	HC (n=73)	*P* value (*t* test/Wilcoxon test)
BMI (kg/m^2), mean (SD)	27.5 (4.9)	27.2 (5.2)	.71
Weight (kg), mean (SD)	78.5 (14.8)	79.8 (15.3)	.55
Height (cm), mean (SD)	169 (7.9)	171 (9.3)	.048
Gender, n (%)			
Men	62 (48)	35 (48)	.95
Women	66 (52)	38 (52)	—[c]
Age class (y), n (%)			
<25	13 (10.2)	7 (11)	—
25-45	60 (46.9)	27 (37)	—
45-65	46 (35.9)	31 (42.5)	—
>65	9 (7)	7 (9.6)	—
Age (y), mean (SD)			
At pump initiation	34 (15)	45 (15)	<.001
At closed-loop initiation	43 (14)	46 (15)	.15
Pump model, n (%)			
Medtronic 780G (with Guardian 4 sensor)	99 (77)	67 (92)	—
“Control IQ” (Tandem Slim 2X pump, Dexcom G6 sensor)	17 (9)	6 (8)	—
“CamAPS” (Ypsopump, Dexcom G6 sensor)	12 (13)	—	—
Baseline glucose control, mean (SD)			
TIR[d] (%)	52.9 (16)	65.9 (15.1)	<.001[e]
GRI[f]	56.4 (21)	37.8 (19.8)	<.001[e]

[a]CIRDIA: Inter-Regional Center for Automated Insulin in Diabetes.

[b]HC: hospital center.

[c]Not applicable.

[d]TIR: time in range 70-180 mg/dL.

[e]Wilcoxon test values.

[f]GRI: glycemia risk index.

XSL•FO
RenderX

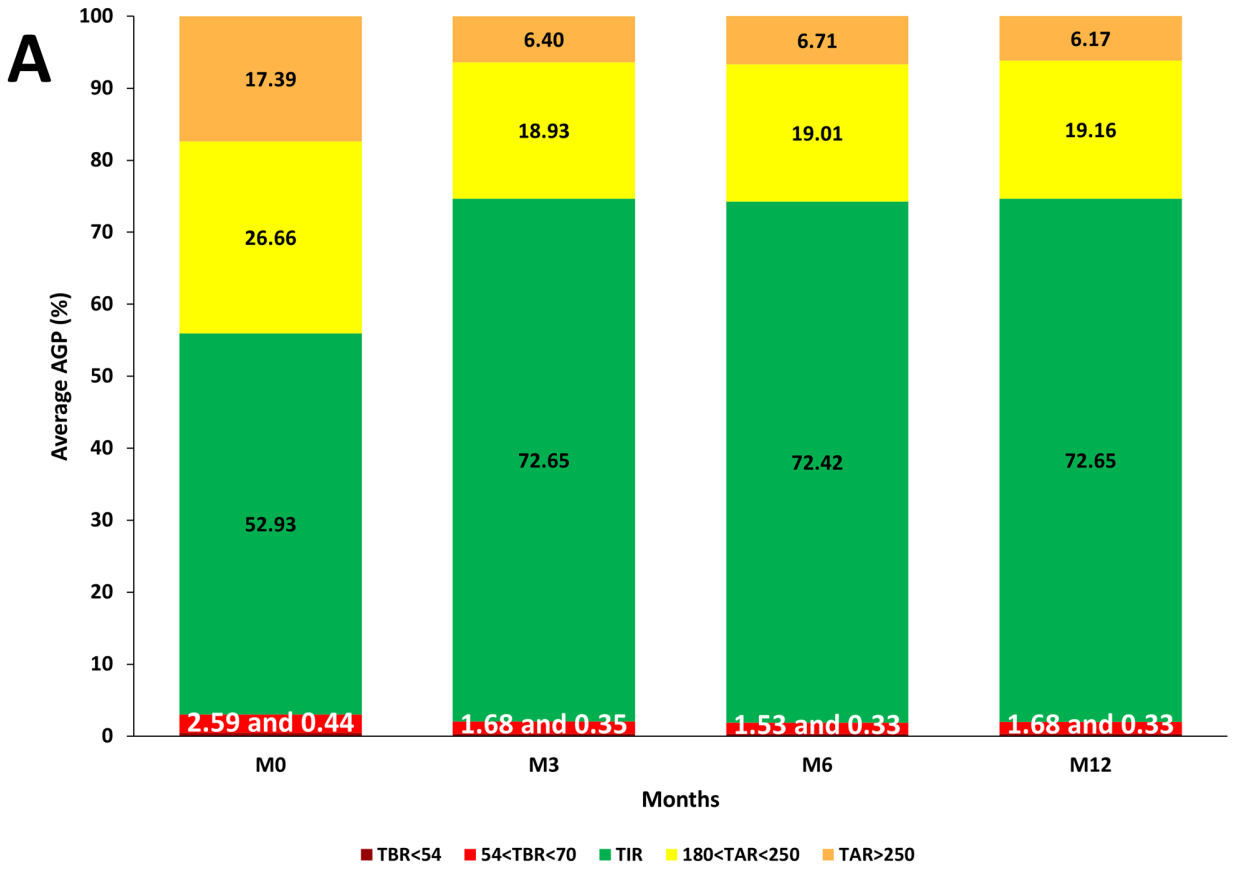
Efficacy and Safety Outcomes

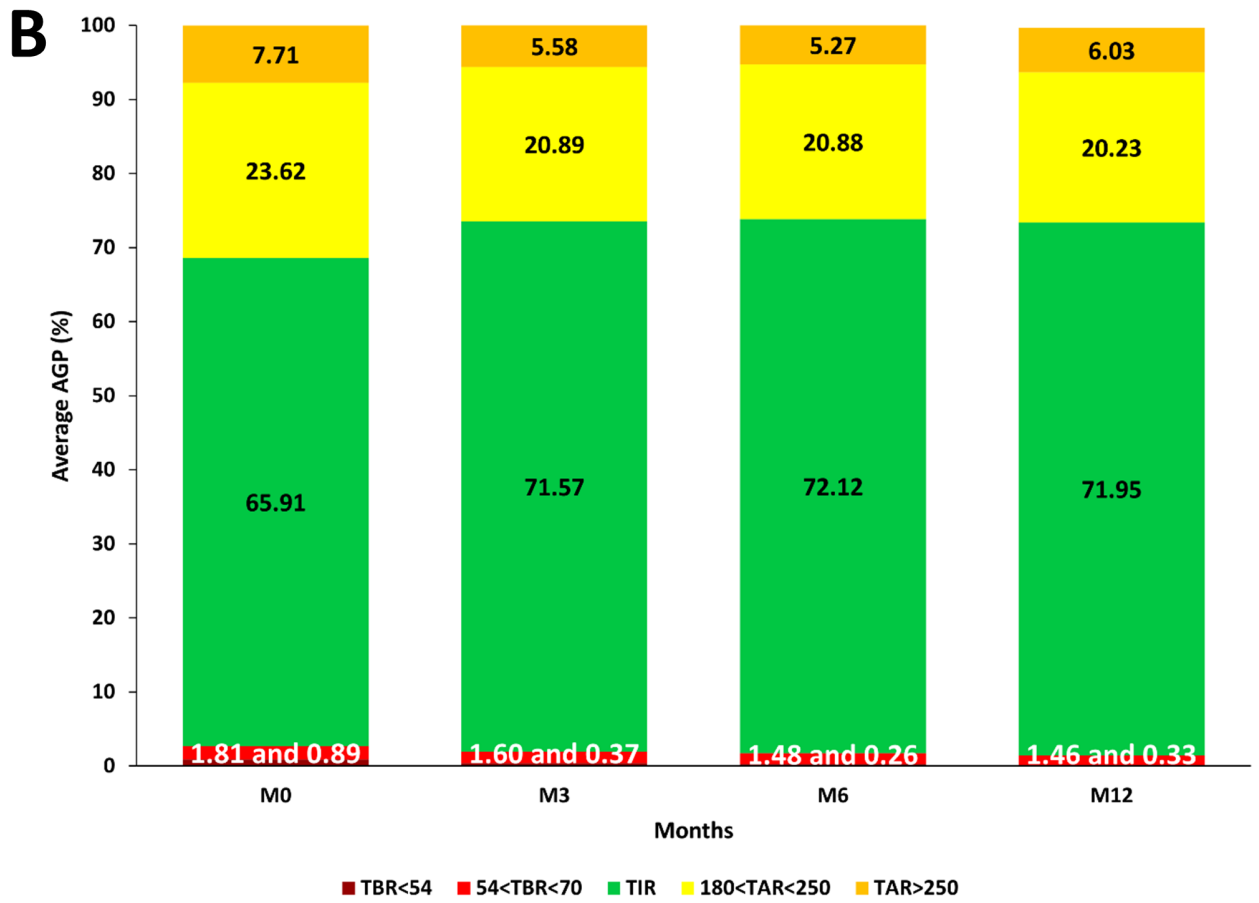
Figure 1 illustrates the changes in the ambulatory glucose profile from baseline to 1 year after initiation.

XSL•FO
RenderX

Figure 1. Ambulatory glucose profile for both comparison arms (A: CIRDIA group, B: hospital centers group) at baseline (M0) and after 3 months (M3), 6 months (M6), and 12 months (M12) of closed-loop use. AGP: ambulatory glucose profile; CIRDIA: Inter-Regional Center for Automated Insulin in Diabetes; TAR: time above range; TBR: time below range; TIR: time in range.

XSL•FO
RenderX





XSL•FO
RenderX

Figure 2 presents GRI grids showing glycemic risk zones over the same period (zone A: minimal hypo- or hyperglycemia risk; zone E: maximal hypo- or hyperglycemia risk). At baseline, 79% (101/128) of the patients from the CIRDIA group were in the intermediate risk (zone C) or high-risk zones (zones D and E). After 1 year on CL, only 21% (27/128) remained in these GRI zones. In the HC arm, 34% (25/73) of the patients were in zones C, D, and E at baseline, and 25% (18/73) remained in these zones after 1 year.

Figure 2. Glycemia risk index (GRI) grids at baseline (M0) and 1 year after closed-loop initiation (M12). Upper grids: Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) group; lower grids: hospital center (HC) group. Each participant is identified by a blue circle and their identification number.

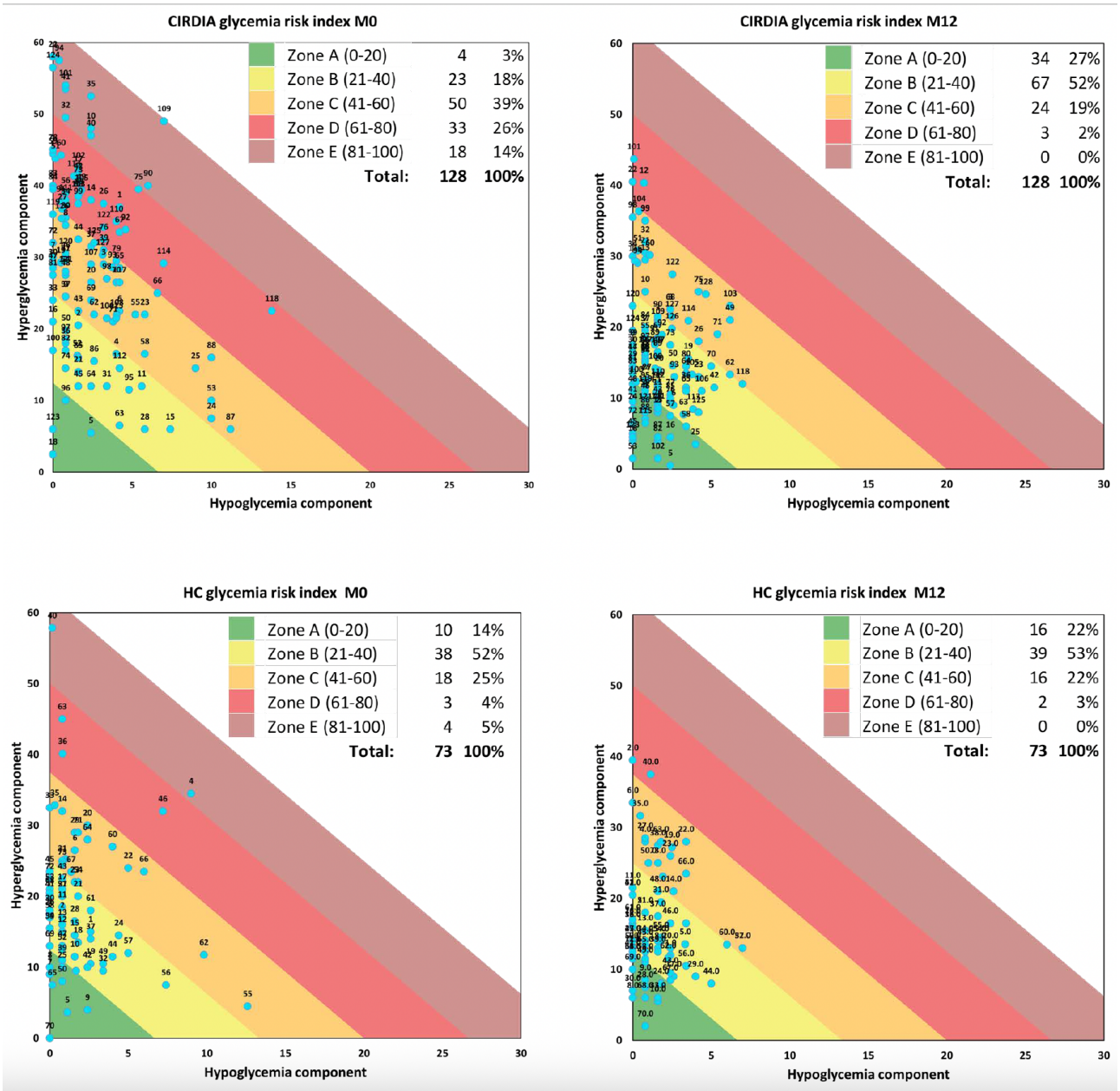


Table 2 summarizes the overall effectiveness and safety results for the total population and according to age at inclusion. After 12 months, the mean (SD) TIR increased by 19.8 points in the CIRDIA group (from 52.9% [16] to 72.7% [11.6]) and by 6 points in the HC group (from 65.9% [15.1] to 71.9% [10.5]). Although baseline differences were significant (P<.001), no significant difference between groups was observed at 12 months (P=.60). The GRI decreased in both groups, by 26.3 points in the CIRDIA group, from 56.4 (21) to 30.1 (14.1), and by 7.5 points in the HC arm, from 37.8 (19.8) to 30.3 (13). No significant difference between groups was observed at 12 months (P=.91).

XSL•FO
RenderX

Table . Glycemic outcomes at baseline (M0) and 12 months (M12) after closed-loop initiation for the total population and according to age class at inclusion[a].

Parameters	CIRDIA[b] (n=128), mean (SD)	HC[c] (n=73), mean (SD)	*P* value
M0[d]			
TIR[e]			
Total	52.9 (16)	65.9 (15.1)	<.001
Age class (y)			
<25	50.4 (11.4)	67.8 (14.7)	.01
25-45	48.4 (15.6)	64.7 (16.7)	<.001
45-65	57.6 (16.3)	65.8 (14)	.02
≥65	62.9 (14.2)	68.9 (14.2)	.30
GRI[f]			
Total	56.4 (21)	37.8 (19.8)	<.001
Age class (y)			
<25	59.6 (16.9)	35.3 (17.9)	.007
25-45	63 (20.7)	40.2 (22.1)	<.001
45-65	50.2 (20.1)	36.6 (17.7)	.002
≥65	39.3 (16.6)	36.7 (24.3)	.70
M12[g]			
TIR			
Total	72.7 (11.6)	71.9 (10.5)	.60
Age class (y)			
<25	70 (11.8)	77.9 (11.8)	.20
25-45	69.1 (11.7)	72 (7.9)	.20
45-65	76.7 (10.6)	72.2 (10.8)	.09
≥65	79 (6.2)	63.6 (14)	.03
GRI			
Total	30.1 (14.1)	30.3 (13)	.91
Age class (y)			
<25	33.6 (14.5)	23.5 (15.1)	.10
25-45	34.4 (13.9)	29.7 (10.7)	.12
45-65	25.3 (13.1)	30.2 (13)	.13
≥65	21.2 (7.4)	40.9 (15.3)	.01

[a]Values were compared using the Wilcoxon rank sum test.

[b]CIRDIA: Inter-Regional Center for Automated Insulin in Diabetes.

[c]HC: hospital center.

[d]M0: closed-loop initiation.

[e]TIR: time in range 70 - 180 mg/dL.

[f]GRI: glycemia risk index.

[g]M12: 12 months after closed-loop initiation.

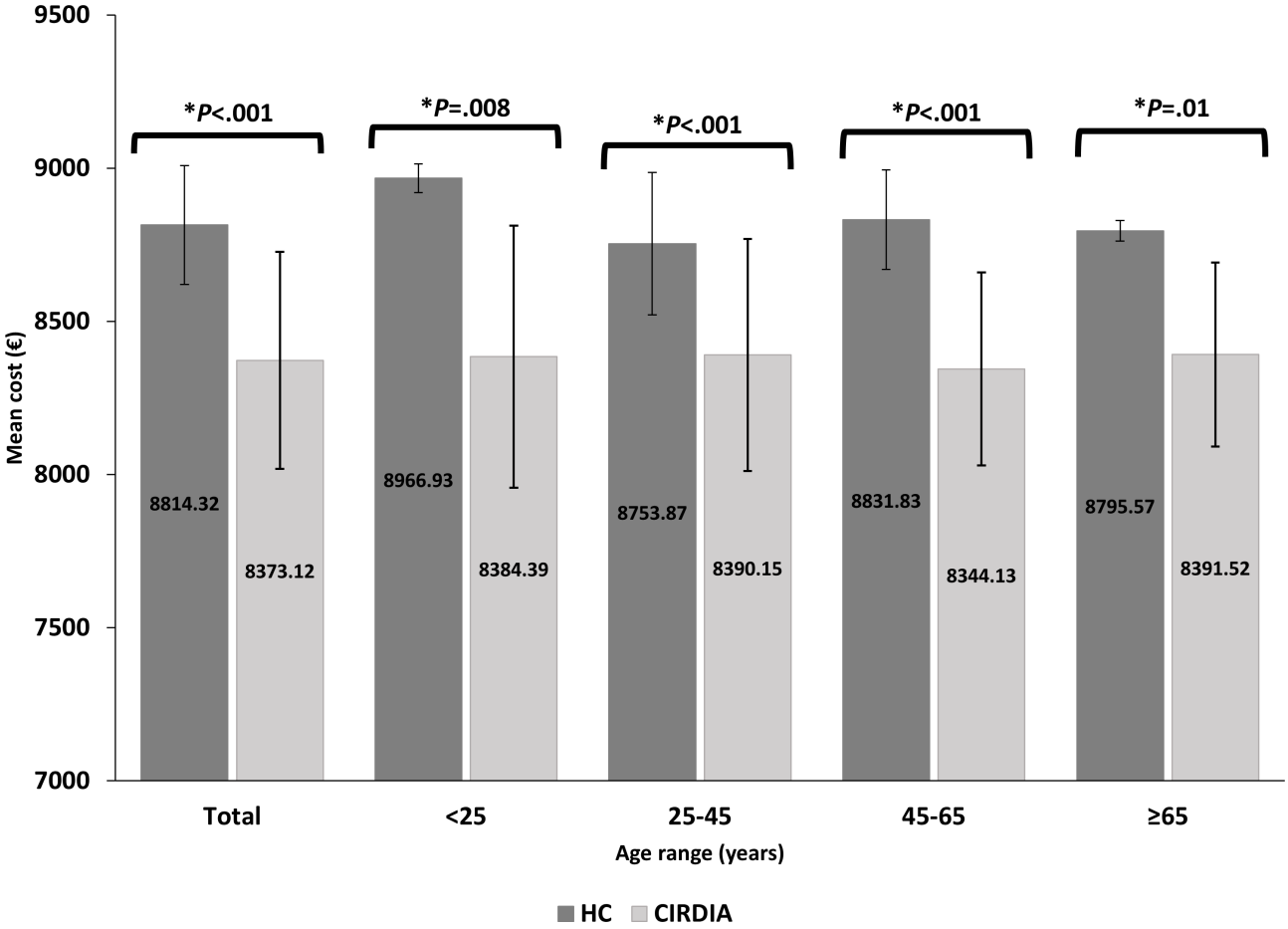
Subgroup analyses revealed no statistically significant differences between the CIRDIA and HC at M12, except among patients older than 65 years, for whom CIRDIA participants had higher TIR and lower GRI values (*P*=.03 and *P*=.01, respectively).

XSL•FO
RenderX

Costs Outcomes

We combined all cost items by type of procedure, year, and data source. Costs are expressed in euros from the French National Health Insurance perspective, and Figure 3 shows the mean costs for both comparison arms and by subgroup.

Figure 3. Average costs per patient according to the care setting. €1=US $1.10 at the time of the study. CIRDIA: Inter-Regional Center for Automated Insulin in Diabetes; HC: hospital centers.



The total cost of CL insulin therapy management was €1,077,231 (1 €=1.10 US $ at the time of the study) for 128 patients initiated in the CIRDIA, which was a mean cost of €8373.12 (SD 427.3) per patient. In the HC group, the total cost was €645,991 for 73 patients, which was a mean cost of €8814.32 (SD 192) per patient. Out-of-hospital–based management was associated with significantly lower costs (P<.001). All cost components are shown in Table S1 in Multimedia Appendix 1.

Incremental Cost-Effectiveness Ratio

The base-case analysis, using mean parameter values, indicated that the CIRDIA was less costly while achieving comparable effectiveness and safety to HC. This situation corresponds to dominance, with an estimated saving of €626 per additional percentage point of TIR and €2011 per point reduction in GRI, indicating that CL initiation in HC is dominated by the CIRDIA. The detailed results are presented in Table 3.

XSL•FO
RenderX

Table . Base-case cost-effectiveness and cost-safety results.

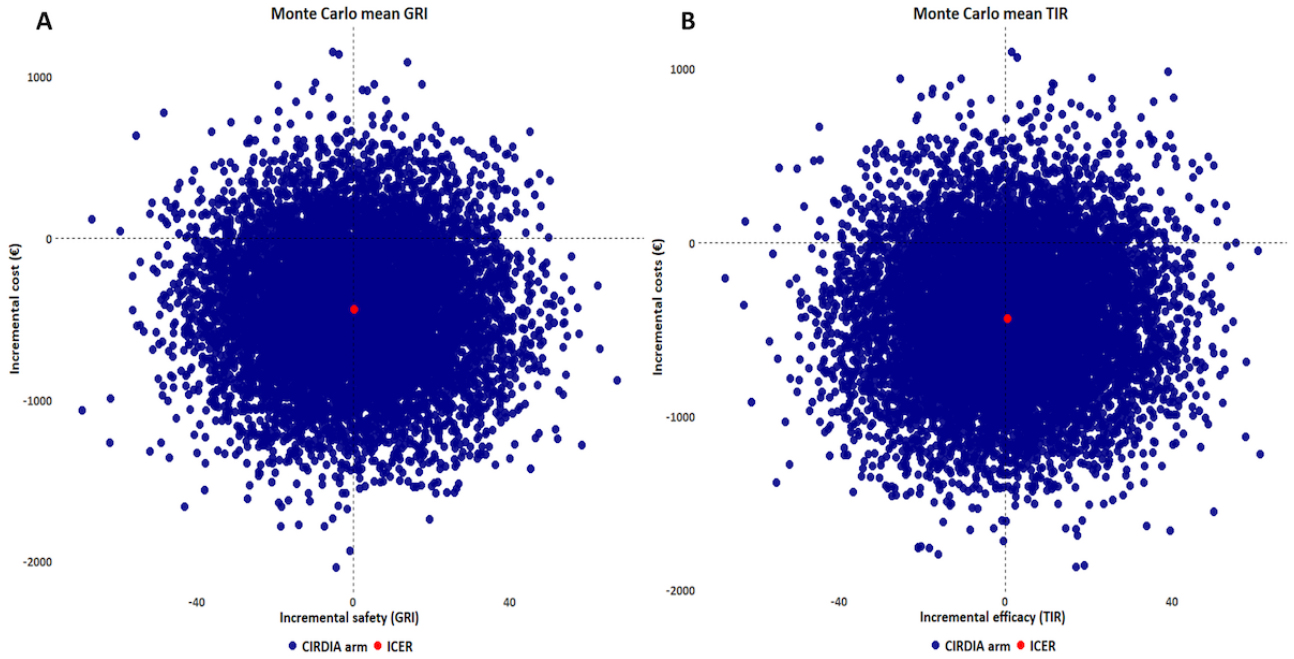
Parameters	CIRDIA[a]	HC[b]
Costs per patient (€[c])	8373.12	8814.32
Incremental costs (€)	−441.20	N/A[d]
Mean efficacy (TIR[e])	72.65	71.95
Incremental efficacy (increase in TIR)	0.70	N/A
Mean safety (reduction in GRI[f])	30.11	30.33
Incremental safety	−0.22	N/A
ICER[g]	−625.83	N/A
ICSR[h]	2011.02	N/A

[a]CIRDIA: Inter-Regional Center for Automated Insulin in Diabetes.
[b]HC: hospital center.
[c]€1=US $1.10 at the time of the study.
[d]N/A: not applicable.
[e]TRI: time in range.
[f]GRI: glycemia risk index.
[g]ICER: incremental cost-effectiveness ratio (based on the increase of time in range).
[h]ICSR: incremental cost-safety ratio (based on the decreased of glycemia risk index).

Sensitivity Analysis

In the PSA (10,000 simulations), the CIRDIA was less costly in 8600 (86%) of the cases. Strong dominance (less costly and more effective) was observed in 4340 (43.4%) of the simulations, while in 4270 (42.7%) of the simulations, the CIRDIA was less costly but less effective. The probability of being more or less effective was generally consistent with the base-case results. Only 1400 (14%) of the simulations placed the CIRDIA in a more costly position, being either less effective (n=700, 7.0%) or more effective (n=690, 6.9%). The scatter plot of the incremental cost-effectiveness plane is presented in Figure 4.

Figure 4. Probabilistic sensitivity analysis of 10,000 Monte Carlo simulations using glycemia risk index (GRI) (A) or time in range (TIR) (B) Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA). €1=US $1.10 at the time of the study. ICER: incremental cost-effectiveness ratio.



XSL•FO
RenderX

Discussion

Principal Findings

The aim of this study was to assess the cost-effectiveness of CL initiation by the CIRDIA and HC from a health care payer perspective. We examined the relationship between initiation models and the increase in TIR or reduction in GRI and whether or not an out-of-hospital CL initiation and follow-up can be achieved in a cost-effective manner compared to the usual hospital management. In our cohort, CL initiation through the CIRDIA was associated with comparable TIR and GRI values at 12 months compared to HC initiation (P=.60 and P=.91, respectively), while being consistently less costly (P<.001), although baseline TIR was lower and baseline GRI was higher in the CIRDIA group. Sensitivity analyses further supported these results, confirming that the CIRDIA generally remained less costly than HC across a wide range of parameter variations.

Prior Work

To our knowledge, this is the first cost-effectiveness analysis comparing hospital-based and out-of-hospital CL initiation. However, our findings are consistent with previous studies, such as Böhme et al [29], which reported no significant differences in effectiveness between outpatient care and hospital settings in therapeutic education programs for patients with type 2 diabetes in France. Similarly, Cavassini et al [30] reported that the outpatient management of gestational diabetes was more cost-beneficial than hospital-based care in Brazil, underlining the potential economic advantages of ambulatory strategies. In the United Kingdom, Pulleyblank et al [31] also found that treatment setting had a significant impact on costs in patients with type 2 diabetes, with outpatient follow-up being less resource-intensive than hospital-based management.

Moreover, recent studies have shown that transitioning to CL reduces the GRI at 1 year [32-34], which is consistent with the trend observed in our cohort.

Strength

One major strength of this study is the use of real-world French data, but many published economic evaluations of CL systems have so far relied mainly on modeled analyses conducted in the United States and the United Kingdom. Furthermore, the use of TIR and GRI as primary end points is relatively novel in economic evaluations, allowing for the integration of a clinically relevant weighting of risk in the assessment of glycemic control [19,35].

Finally, sensitivity analyses and subgroup explorations provided additional insights into the robustness of our results, supporting the finding that CIRDIA and HC achieved broadly comparable outcomes in terms of TIR and GRI, whereas at baseline, TIR was lower and GRI higher in the CIRDIA participants. This also underlines that prior to CL initiation, patients followed in out-of-hospital settings do not have better glucose control than those followed in hospital centers, at least in our population.

Limitations

This study has several limitations.

First, the relatively small sample size limits the representativeness of the study population and, consequently, the robustness of the conclusions.

Second, there was an imbalance in baseline efficacy and safety outcomes between groups, which could have led to selection bias. To address this, we performed inverse probability of treatment weighting to adjust for sociodemographic characteristics as well as baseline efficacy and safety measures. After weighting, the cost advantage of the CIRDIA was maintained, and the results on effectiveness and safety suggested a potential benefit, although these should be interpreted cautiously given the limited sample size (data not shown). However, the patients who chose to start CL therapy in the CIRDIA setting might be different from the patients from the HC group in terms of prior education or other characteristics. A prospective study with better characterizations of these items will be needed.

Third, the 1-year time horizon restricts the evaluation to the short term and does not allow assessment of long-term effectiveness or costs, although this choice was justified by the specific objective of analyzing the initiation phase of CL.

Fourth, because the costs were assessed using French Health Insurance (Assurance Maladie) rates, the results may not be generalizable to other health care systems. However, this study suggests that CL initiation in an outpatient setting is feasible, safe, and probably less expensive than the inpatient setting, regardless of the health care system.

Fifth, we cannot exclude a bias in the recruitment of HC patients as it is possible that the patients sent to Santelys home health care provider by the hospital teams might have a different (here better) control compared to other HC patients. However, as patients are from 5 different hospitals, it is unlikely that this happened in all of the hospitals.

Finally, missing information on complications and comorbidities in the hospital arm may have led to an underestimation of certain costs (eg, retinopathy-related tests), although this does not appear to alter the overall trend observed.

Nevertheless, the data from the French Closed-Loop Observatory (OB2F) indicate that outpatient initiation is already widespread, reinforcing the relevance of investigating this organizational model [18].

Conclusion

This cost-effectiveness analysis compared 2 models of CL initiation for patients with T1D: a conventional hospital-based model and an out-of-hospital–based model supported by the CIRDIA.

Although baseline TIR was lower and baseline GRI was higher in the CIRDIA out-of-hospital setting compared to the HC setting, our results showed no significant differences in efficacy or safety outcomes between the 2 approaches. However, the CIRDIA setting was associated with lower management costs. While the patients who choose to initiate a CL system in the CIRDIA setting are probably not the same as the patients who choose to initiate CL in hospitals, these real-life findings suggest that the CIRDIA may represent a viable alternative

XSL•FO
RenderX

organizational model for CL initiation in France, as it combines efficacy and savings.

Future research should assess whether these results hold over longer time horizons (eg, 5 or even 10 y) and from broader perspectives, such as a societal perspective that incorporates quality of life and indirect costs. Such work would enable cost-utility analyses to complement our cost-effectiveness findings.

From a policy perspective, the resources saved through out-of-hospital CL initiation could be reallocated to organizations such as the CIRDIA, which bring together highly trained diabetologists and uphold high-quality standards. This would allow persons living with T1D to choose their CL initiation setting, ensure early access to new technologies, and benefit the overall health care system through a cost-effective model.

Acknowledgments

We are grateful to (by alphabetical order) Dr Fabienne-Amiot-Chapoutot (Inter-Regional Center for Automated Insulin in Diabetes [CIRDIA], Troyes, France), Dr Véronique Averous (Douai Hospital), Dr Blandine Courbebaisse (CIRDIA, Fontenay-sous-Bois, France), Dr Vincent Degros (Valenciennes Hospital), Dr Joëlle Dupont (CIRDIA and HPL, St Etienne, France), Dr Salha Fendri (Amiens University Hospital), Dr Emmanuelle Lecornet-Sokol (CIRDIA and IPE, Paris, France), Dr Christine Lemaire (Bethune hospital), Dr Estelle Personeni (CIRDIA, Chatenoy-le-Royal, France), Dr Sylvie Picard (CIRDIA, Dijon, France), Pr Anne Vambergue (Lille University Hospital), and Dr Rihab Caillier (Santélys, Loos, France) for their assistance with the recruitment phase and Imane Boukhers, PhD (Santelys, Loos, France), for the editorial support. This study would not have been possible without the support of the CIRDIA.

Funding

This work was supported by a research grant from Santélys, Loos, France. Outside of this work, the CIRDIA has also received a research grant form the Clement-Drevon Foundation, Dijon, France; an award from the FENAREDIAM; and donations from the French Diabetic Association Burgundy-Franche-Comté and Côte d'Or and from the French Mutualité Jura.

Data Availability

The data used in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Authors' Contributions

Conceptualization: FG, SB, SP, TF
Data curation: AR, FB, MN, SP
Formal analysis: AR, MN, TF
Funding acquisition: SP
Investigation: FB, SP
Methodology: AR, FB, SB, SL, TF
Project administration: SP
Supervision: AR, SB, TF
Validation: AR, FB, FG, SB, SL, TF
Writing – original draft: MN
Writing – review and editing: SB, SL, SP, TF

Conflicts of Interest

None declared related to this study. Outside of this work, SP has received consulting and/or speaking fees from Abbott, Air Liquide, Asdia, Dexcom, Insulet, Isis Diabete, Lilly, Medtronic, NHC, Novo Nordisk, Orkyn, Sanofi, and VitalAire. SL has received consulting and/or speaking fees from Abbott, Insulet, Medtronic, and Dexcom. SB has received consulting and/or speaking fees from Dexcom, Insulet, Lilly, and Medtrum.

Multimedia Appendix 1
Costs components for closed-loop system management.
[DOCX File, 18 KB - diabetes_v11i1e86690_app1.docx]

References

1. Diabetes. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2026-01-14]
2. People with diabetes, in 1,000s. International Diabetes Federation. URL: https://diabetesatlas.org/data-by-indicator/diabetes-estimates-20-79-y/people-with-diabetes-in-1000s/ [accessed 2026-01-14]

XSL•FO
RenderX

3. Jarraud C, Piotto E, Leclerc S, Thomassin C, Galmiche H. Le Diabète en France: nouvelles données épidémiologiques [Article in French]. J Epidemiol Popul Health 2025 Mar;73:202923. [doi: 10.1016/j.jeph.2025.202923]
4. Franc S, Schaepelynck P, Tubiana-Rufi N, et al. Mise en place de l'insulinothérapie automatisée en boucle fermée: position d'experts français [Article in French]. Med Mal Metab 2020 Sep;14(5S):S1-S40. [doi: 10.1016/S1957-2557(20)30003-1]
5. Renard E. L'insulinothérapie dans le diabète de type 1: un paysage en constante évolution [Article in French]. Med Mal Metab 2023 Feb;17(1):58-63. [doi: 10.1016/j.mmm.2022.12.001]
6. Benhamou PY, Lablanche S, Vambergue A, et al. The beneficial effects of closed-loop insulin delivery in patients with highly unstable type 1 diabetes eligible for islet transplantation are maintained over 6 months: an extension study of the DBLHU-WP10 trial. Diabetes Obes Metab 2022 May;24(5):956-961. [doi: 10.1111/dom.14654] [Medline: 35099098]
7. Murphy HR, Kumareswaran K, Elleri D, et al. Safety and efficacy of 24-h closed-loop insulin delivery in well-controlled pregnant women with type 1 diabetes: a randomized crossover case series. Diabetes Care 2011 Dec;34(12):2527-2529. [doi: 10.2337/dc11-1430] [Medline: 22011408]
8. Wang LR, Malcolm J, Arnaout A, Humphrey-Murto S, LaDonna KA. Real-world patient experience of long-term hybrid closed-loop insulin pump use. Can J Diabetes 2021 Dec;45(8):750-756. [doi: 10.1016/j.jcjd.2021.02.006] [Medline: 33958309]
9. Choudhary P, Arrieta A, van den Heuvel T, Castañeda J, Smaniotto V, Cohen O. Celebrating the data from 100,000 real-world users of the MiniMed. Diabetes Technol Ther 2024 Mar;26(S3):32-37. [doi: 10.1089/dia.2023.0433] [Medline: 38377326]
10. New data highlights the clinical impact of the MiniMed. Medtronic. 2024 Nov 18. URL: https://news.medtronic.com/New-data-highlights-the-clinical-impact-of-the-MiniMed-TM-780G-system-in-Asia-Pacific [accessed 2026-01-14]
11. Le remboursement du 1er système de boucle fermée hybride: retour sur un long chemin pour l'amélioration de la qualité de vie des personnes diabétiques [Article in French]. Fédération Française des Diabétiques. 2021 Nov 18. URL: https://www.federationdesdiabetiques.org/federation/actualites/le-remboursement-du-1er-systeme-de-boucle-fermee-hybride-retour-sur-un-long-chemin-pour-lamelioration-de-la-qualite-de-vie-des-personnes-diabetiques [accessed 2026-01-14]
12. Raffier N. Le pancréas artificiel allège le fardeau du diabète [Article in French]. Medscape. 2023 May 3. URL: https://francais.medscape.com/voirarticle/3610024 [accessed 2026-01-14]
13. Picard S, Personeni E, Dupont J, et al. Près de la moitié des personnes vivant avec un diabète de type 1 atteignent les objectifs internationaux de glucose 6 mois après l'initiation d'une boucle fermée hybride hors hôpital: résultats du CIRDIA [Article in French]. Ann Endocrinol (Paris) 2024 Oct;85(5):371-372. [doi: 10.1016/j.ando.2024.08.075]
14. Choix méthodologiques pour l'évaluation économique à la HAS [Report in French]. : Haute Autorité de Santé; 2020 Jul 2 URL: https://www.has-sante.fr/upload/docs/application/pdf/2020-07/guide_methodologique_evaluation_economique_has_2020_vf.pdf [accessed 2026-01-14]
15. Le règlement général sur la protection des données - RGPD [Article in French]. CNIL - Commission Nationale de l'Informatique et des Libertés. URL: https://www.cnil.fr/fr/reglement-europeen-protection-donnees [accessed 2026-01-14]
16. Recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé [Report in French]. : Commission Nationale de l'Informatique et des Libertés; 2025 URL: https://www.cnil.fr/sites/default/files/2024-01/mr-004.pdf [accessed 2026-01-14]
17. Picard S, Lecornet-Sokol E. Le CIRDIA, un nouveau concept d'organisation des soins pour le diabète de type 1 en France [Article in French]. Med Mal Metab 2025 May;19(3):235-241. [doi: 10.1016/j.mmm.2025.01.013]
18. Raffier N. SFD 2024 – en « vie réelle », le pancréas artificiel fait ses preuves ! [Article in French]. Univadis. 2024 Apr 8. URL: https://www.univadis.fr/viewarticle/sfd-2024-vie-r%25C3%25A9elle-pancr%25C3%25A9-artificiel-fait-ses-2024a10006ea [accessed 2026-01-14]
19. Raffier N. TITR et GRI: les critères mesure continue du glucose de demain? [Article in French]. Medscape. 2025 Apr 25. URL: https://francais.medscape.com/voirarticle/3612807 [accessed 2026-01-14]
20. ALD n° 8 - diabète de type 1 chez l'adulte [Report in French]. : Haute Autorité de Santé; 2014 Apr 24 URL: https://www.has-sante.fr/jcms/c_419399/fr/ald-n-8-diabete-de-type-1-chez-l-adulte [accessed 2026-01-14]
21. Mallone R, Bismuth E, Thivolet C, et al. Dépistage et prise en charge du diabète de type 1 préclinique, stade 1–2. Prise de position d'experts français [Article in French]. Med Mal Metab 2024 Sep;18(5):405-432. [doi: 10.1016/j.mmm.2024.06.003]
22. Tarifs conventionnels des médecins généralistes et spécialistes [Article in French]. L'Assurance Maladie. 2026 Jan 15. URL: https://www.ameli.fr/roubaix-tourcoing/medecin/exercice-liberal/facturation-remuneration/consultations-actes/tarifs/tarifs-conventionnels-medecins-generalistes-specialistes [accessed 2026-01-14]
23. CCAM en ligne [Article in French]. L'Assurance Maladie. URL: https://www.ameli.fr/accueil-de-la-ccam/index.php [accessed 2026-01-14]
24. Téléchargement de la TNB version 101 [Article in French]. L'Assurance Maladie. URL: http://www.codage.ext.cnamts.fr/codif/nabm/telecharge/index_tele.php?p_site=AMELI [accessed 2025-09-8]
25. Liste des produits et prestations [Article in French]. L'Assurance Maladie. URL: http://www.codage.ext.cnamts.fr/codif/tips/index_presentation.php?p_site=AMELI [accessed 2026-01-14]

XSL•FO
RenderX

26. Tarifs MCO et HAD [Article in French]. Agence technique de l'information sur l'hospitalisation. URL: https://www.atih.sante.fr/tarifs-mco-et-had [accessed 2026-01-14]
27. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Value Health 2022 Jan;25(1):3-9. [doi: 10.1016/j.jval.2021.11.1351] [Medline: 35031096]
28. Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–6. Value Health 2012;15(6):835-842. [doi: 10.1016/j.jval.2012.04.014] [Medline: 22999133]
29. Böhme P, Durain-Sieffert D, Fay R, et al. P89 Efficacité à court terme d'un réseau de ville centré sur l'éducation thérapeutique du patient diabétique de type 2: comparaison avec un service hospitalo-universitaire [Article in French]. Diabetes Metab 2010 Mar;36:A60. [doi: 10.1016/S1262-3636(10)70237-8]
30. Cavassini ACM, Lima SAM, Calderon IMP, Rudge MVC. Cost-benefit of hospitalization compared with outpatient care for pregnant women with pregestational and gestational diabetes or with mild hyperglycemia, in Brazil. Sao Paulo Med J 2012;130(1):17-26. [doi: 10.1590/s1516-31802012000100004] [Medline: 22344355]
31. Pulleyblank R, Laudicella M, Olsen KR. Cost and quality impacts of treatment setting for type 2 diabetes patients with moderate disease severity: hospital- vs. GP-based monitoring. Health Policy 2021 Jun;125(6):760-767. [doi: 10.1016/j.healthpol.2021.03.009] [Medline: 33789816]
32. Picard S, Courbebaisse B, Dupont J, et al. Glycemia Risk Index (GRI) and international glucose targets before and 6 months after initiation of hybrid closed loop system in the CIRDIA, a French multisite out-of-hospital center. Diabetes Metab 2025 Mar;51(2):101617. [doi: 10.1016/j.diabet.2025.101617] [Medline: 39922278]
33. Resmini E, Zarra E, Dotti S, et al. Impact on Glycemia Risk Index and other metrics in type 1 adult patients switching to Advanced Hybrid Closed-Loop systems: a one-year real-life experience. Eur J Med Res 2024 Jul 15;29(1):365. [doi: 10.1186/s40001-024-01946-w] [Medline: 39004734]
34. Castorani V, Rigamonti A, Frontino G, et al. Turning the tides: achieving rapid and safe glucose control in adolescents with suboptimally controlled type 1 diabetes using advanced hybrid closed loop systems. Front Endocrinol (Lausanne) 2024;15:1243565. [doi: 10.3389/fendo.2024.1243565] [Medline: 38628580]
35. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023 Jan;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9] [Medline: 36493795]

Abbreviations

CGM: continuous glucose monitoring
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CIRDIA: Inter-Regional Center for Automated Insulin Therapy
CL: closed-loop
DH: day hospitalization
GRI: glycemia risk index
HAS: French National Health Authority
HC: hospital center
ICER: incremental cost-effectiveness ratio
PSA: probabilistic sensitivity analysis
SFD: French-Speaking Diabetes Society
T1D: type 1 diabetes
TIR: time in range

Edited by A Stone; submitted 28.Oct.2025; peer-reviewed by B Detournay, J Lu; revised version received 28.Dec.2025; accepted 07.Jan.2026; published 29.Jan.2026.

Please cite as:
Napame M, Picard S, Foglia T, Rubenstrunk A, Baudoux F, Giroud F, Lablanche S, Borot S
Inter-Regional Center for Automated Insulin in Diabetes (CIRDIA) and Hospital-Based Approaches to Closed-Loop Therapy in Type 1 Diabetes: Cost-Effectiveness Analysis
JMIR Diabetes 2026;11:e86690
URL: https://diabetes.jmir.org/2026/1/e86690
doi:10.2196/86690

XSL•FO
RenderX

© Mercia Napame, Sylvie Picard, Tony Foglia, Anne Rubenstrunk, Florence Baudoux, Fancoise Giroud, Sandrine Lablanche, Sophie Borot. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Use of the GTT@home Oral Glucose Tolerance Test Kit in Gestational Diabetes Mellitus: Performance Evaluation Study

Gareth J Dunseath[1], PhD; Stephen D Luzio[1], PhD; Wai Yee Cheung[1], PhD; Sharon N Parsons[1], PhD; Nicola John[2]; Mahmoud Chokor[2], Grad Dip, MRCP; Michael Atkinson[2], MBBCh, MRCP; Rajesh Peter[2], MD, FRCP

[1]Diabetes Research Group, Swansea University, Ground Floor, Grove Building, Singleton Campus, Swansea, United Kingdom

[2]Neath Port Talbot Hospital, Port Talbot, United Kingdom

Corresponding Author:
Gareth J Dunseath, PhD
Diabetes Research Group, Swansea University, Ground Floor, Grove Building, Singleton Campus, Swansea, United Kingdom

Abstract

Background: The 75-g oral glucose tolerance test (OGTT) remains the optimal diagnostic test for use in pregnancy but needs to be performed in the clinical setting. The GTT@home OGTT device offers the potential to enable patients to perform the test at home using capillary blood samples.

Objective: This study aimed to determine the accuracy of the GTT@home device compared to the routine National Health Service laboratory reference method using blood samples during an OGTT from pregnant women at high risk of developing gestational diabetes mellitus (GDM).

Methods: A total of 65 women (aged >18 y), at high risk for GDM (per the National Institute for Health and Care Excellence guidelines) were recruited for this performance evaluation. Following an overnight fast, participants went for a 75-g OGTT. Fasting and 2-hour capillary glucose levels were measured using the GTT@home device with corresponding venous samples measured in the laboratory.

Results: The complete data for analysis was available for 61/65 devices. The overall bias for the GTT@home device was +0.16 mmol/L. Correlation analysis of the clinical performance of the two methods using a surveillance error grid showed 79.8% of results in the lowest, 16.9% in the “slight, lower” and 2.4% in the “slight, higher” risk categories. Only 0.8% were “moderate, lower” risk, and none were in any higher risk categories. There was agreement in the classification in 54/61 cases. The GTT@home device under-classified 2 cases and over-classified 5 cases.

Conclusions: The GTT@home device worked well in a controlled, antenatal clinical setting. Differences in classification observed were generally due to small differences in glucose values close to the diagnostic cut-offs. The GTT@home device shows promise for home testing of glucose tolerance in pregnant women.

(JMIR Diabetes 2026;11:e69695) doi:10.2196/69695

KEYWORDS

device; diagnosis; gestational diabetes; OGTT; glucose; oral glucose tolerance test

Introduction

Gestational diabetes mellitus (GDM) is a common metabolic disorder occurring in up to 10% of pregnancies in the Western world [1]. Most women with GDM are asymptomatic, and therefore, it is important to screen, diagnose, and manage the condition, as it is associated with an increased risk of maternal and perinatal complications such as pre-eclampsia, macrosomia, shoulder dystocia, and neonatal hypoglycemia. In the United Kingdom, in line with the National Institute for Health and Care Excellence (NICE) guidelines, women with a high risk of GDM are offered a 75-g oral glucose tolerance test (OGTT) at 24 - 28 weeks gestation [2], with plasma glucose levels ≥ 5.6 mmol/l or 2-hour plasma glucose levels ≥ 7.8 mmol/l being diagnostic of GDM.

Unlike diabetes mellitus in the general population, where the 2-hour 75-g OGTT has largely been superseded by the glycosylated hemoglobin level for diagnosis, the OGTT carried out in a clinical setting is still the optimal test for use in pregnancy. The glycosylated hemoglobin level is not sufficiently sensitive to substitute for the OGTT as a screening test, due to the variations in red blood cell turnover seen in pregnancy. However, during the COVID-19 pandemic, changes to diagnosis of GDM were forced on maternity services, with diagnosis being determined by the measurement of either glycosylated hemoglobin or a single fasting or random plasma glucose sample. A retrospective study of prospectively collected data [3] has shown that using the Royal College of Gynecologists COVID-19 Gestational Diabetes screening guideline failed to detect 47 of 82 (57%) women subsequently identified with

XSL•FO
RenderX

GDM and therefore could not be recommended for general use. A scoping review of the guidelines and diagnostic studies evaluating the recommended testing strategies [4] concluded that the OGTT remains the most effective test to identify abnormal glucose tolerance in pregnancy.

In practice, pre-analytical processing of blood samples can affect plasma glucose concentrations due to continuing glycolysis by red cells prior to centrifugation [5-11]. The American Diabetes Association and American Association of Clinical Chemistry recommend that samples for plasma glucose measurement should be collected into sodium fluoride tubes and placed in an ice-water slurry prior to centrifugation within 30min [12,13]. If a delay in centrifugation is anticipated, citrate tubes should be used as citrate more rapidly inhibits glycolysis [12,13]. In routine clinical practice, this is rarely carried out, which could affect the diagnosis of GDM.

The GTT@home OGTT device is an electronic device that has the potential to enable patients to perform an OGTT from home using capillary blood samples. In non-pregnant women, the GTT@home device has been previously shown to be easy to use, reliable and demonstrated excellent agreement with the results obtained from a reference laboratory analyzer (YSI 2300 stat Plus) [14]. It now needs to be established how results from this device compare with results obtained conventionally from an OGTT in women at risk of GDM. In this study, glucose concentrations during an OGTT were tested with fresh blood samples from women at risk of GDM in the United Kindgom and compared to routine laboratory glucose concentrations.

Methods

Ethical Considerations

Women gave written informed consent to take part in the study. Ethical approval was obtained from Health Research Authority and Health and Care Research Wales, Wales REC 6 (22/WA/0153). Patients or the public were not involved in the design, or conducting, reporting, or dissemination plans of our research. All participant data were anonymized prior to analysis. Participants received no compensation for taking part in this study.

Participants

This performance evaluation study was carried out with women presenting to the antenatal clinic at Neath Port Talbot Hospital for a routine 75-g OGTT at approximately 24 - 28 weeks of gestation. Study participants (n=65) were female, aged >18 years, were at high risk for developing GDM according to NICE guidelines (ie, previous macrosomic baby weighing >4.5 kg or >90th centile, previous GDM, family history of diabetes [first-degree relative with diabetes], or an ethnicity with a high prevalence of diabetes). Participants were excluded if they were unable or unwilling to give informed consent.

GTT@home Device

The Home OGTT devices were provided by Digostics Ltd (Figure 1). The device consists of 2 glucose dehydrogenase test strips (0 and 2-h) with user activated buttons. The test procedure was driven by an integral clock and timer, with audible and visual prompts. Each single-use, disposable device was stored in sealed packaging and opened immediately prior to use. The device was activated by removing a protective cover over the 0-hour test strip (A) and the capillary blood sample being placed on the test strip. Following consumption of the glucose drink, the “set” button is pressed to begin the timer, and after 2 hours, an audible alarm alerts the user to press “stop” and repeat the sampling process with the 2-hour test strip (B). A further audible alarm confirms the test was complete. A detachable data recorder is scanned, and the result automatically transferred to a secure web-based database. No results are visible to the user.

XSL•FO
RenderX

Figure 1. Example of the GTT@home device including user instructions. GTT: glucose tolerance test.

XSL•FO
RenderX

Study Procedure

Participants attended the antenatal clinic at Neath Port Talbot Hospital on one occasion having fasted for 8 hours. At time 0, a finger prick blood sample was collected using a finger pricking device. The blood drop was applied to the 0 minutes (A) glucose sensor on the GTT@home device. A venous blood sample (2 ml) was also taken for routine hospital laboratory analysis as per usual practice (ie, the Roche Cobas enzymatic method). Following the 0-min sample, a drink containing 75 g glucose was given to the subject. At 2 hours following the glucose drink, a second finger prick and venous blood sample were collected. The finger prick capillary blood was applied to the 2-hour (B) sensor on the GTT@home device, and the venous blood was sent to the routine hospital laboratory for blood glucose measurement. Following the 2-hour sample, the participants were allowed to leave the clinic.

The readings from the two sets of tests were generated independently in different locations. Assessors were not able to be influenced by prior knowledge of readings from the other test. Clinical staff collecting the clinical information and performing the GTT@home test were blinded to these results.

The GTT@home device readings were generated automatically and were non-editable. Results were stored electronically on a detachable data chip, scanned, and uploaded to a password-protected database.

Data Analysis

Based on a GDM prevalence of 35%, to achieve the minimally acceptable kappa score of 0.61, a sample size of 65 patients was required to achieve 80% power and a statistical significance of 0.05.

Bias was assessed using Bland-Altman plots of the GTT@home device versus the routine hospital laboratory method [15]. Correlation of the clinical performance of the two methods was assessed using a surveillance error grid [16]. Agreement of the diagnoses generated using the GTT@home device and the routine laboratory analyzer with the reference was assessed using receiver operating characteristic curves (sensitivity and specificity) and positive and negative predictive values [17]. Agreement between the readings was assessed using kappa analysis for the whole OGTT, fasting, and 2-hour glucose values. Patients were categorized as having normal glucose tolerance or being intolerant (0 or 1).

The study classified the status of glucose tolerance according to the NICE criteria [1] with glucose intolerance defined as a fasting plasma glucose level ≥5.6 mmol/l and/or a 2-hour plasma glucose level ≥7.8mmol/l.

Results

The mean (SD) age of participants was 29.6 (4.8; range 19-41) years, and the mean (SD) BMI was 32.8 (7.6; range 16.5-54.7) kg/m^2 (Table 1).

Results from 4 participants were not included in the final analysis: two devices reported only fasting glucose values, one device reported no results, and one patient experienced nausea, so the venous sample was not collected.

The bias of the GTT@home device compared to the routine laboratory method is shown in Figure 2. For fasting plasma glucose, the bias (lower/upper limit of agreement) was 0.01 (−1.13/1.15) mmol/L; at 2 hours, it was +0.31 (-1.84/2.46) mmol/L; and for all results, the bias was +0.16 (−1.57/1.89) mmol/L.

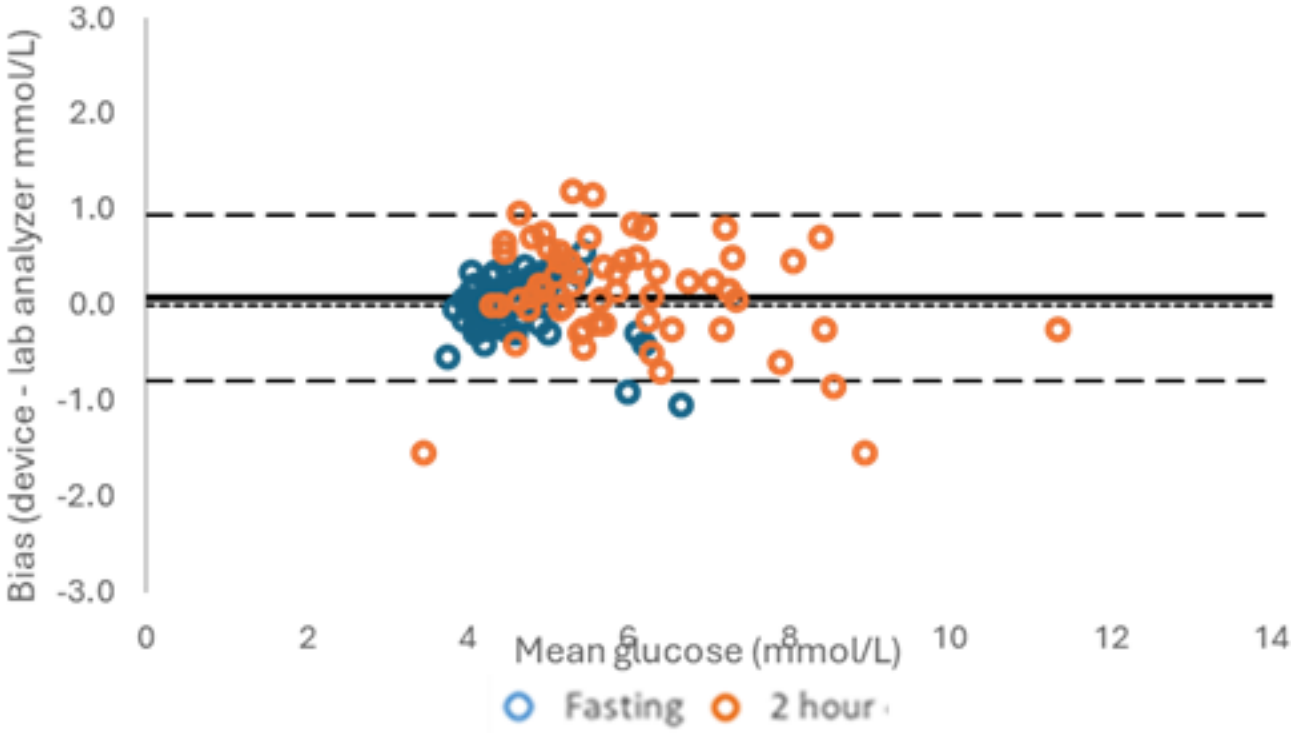
Correlation of the clinical performance of the two methods is shown in Figure 3 using a surveillance error grid. A total of 79.8% of results were in the lowest risk category (“none”), 16.9% in the “slight, lower” risk category, and 2.4% in the “slight, higher” risk category. Only 0.8% were in the “moderate lower” risk category, and there were no results in any of the higher risk categories.

Table . Demographic characteristics.

Demographic characteristic	Mean (SD)	Range
Age (y)	29.6 (4.8)	19.0-41.0
Height (m)	1.64 (0.1)	1.53-1.83
Weight (kg)	89.3 (23.1)	40.0-148.0
BMI (kg/m^2)	32.8 (7.6)	16.6-54.7

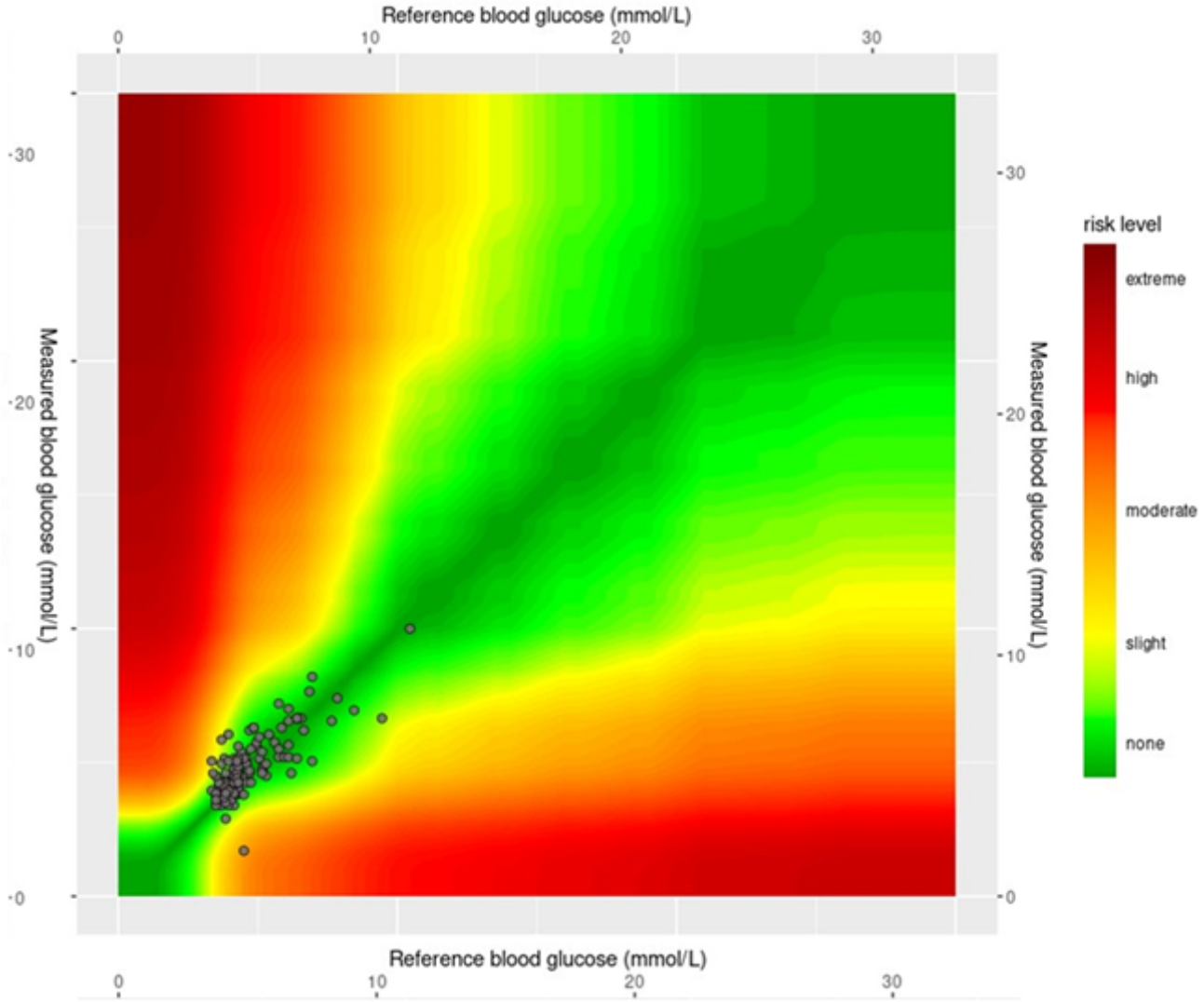
XSL•FO
RenderX

Figure 2. Bias of the GTT@home device. The blue circles indicate fasting levels, while the orange circles indicate the levels after 2 hours.



XSL•FO
RenderX

Figure 3. Surveillance error grid (SEG).



	SEG Risk Level	SEG Risk Category	Number of Pairs	Percent
1	0	None	99	79.8%
2	1	Slight, Lower	21	16.9%
3	2	Slight, Higher	3	2.4%
4	3	Moderate, Lower	1	0.8%
5	4	Moderate, Higher		NA
6	5	Severe, Lower		NA
7	6	Severe, Upper		NA
8	7	Extreme		NA

Cross-tabulation of overall classification (combined fasting and 2 h) is shown in Table 2. There was overall agreement in classification in 54 out of 61 cases. The GTT@home device under classified 2 cases and over classified 5 cases. Of these small number of different classifications, the majority were close to the diagnostic thresholds, with one-third fasting plasma glucose results and two-third 2-hour results within 0.2 mmol/L of the respective cut-offs.

Receiver operating characteristic curves are shown in Figure 4, and positive and negative predictive values in Table 3. The AUC, sensitivity and specificity using receiver operating characteristic analysis for fasting glucose was 0.947, 0.75 and 0.966, respectively, and those for the 2-hour glucose were 0.932, 0.4 and 0.929, respectively. The kappa statistic for the GTT@home device compared to routine laboratory measurement was 0.457 (0.641 for fasting values and 0.301 for

XSL•FO
RenderX

2-h values). The diagnosis of glucose intolerance using both OGTT concentrations showed a positive predictive value of 0.44 and negative predictive value of 0.96.

Table . Cross tabulation.

		GTT@home device		Total
		NGT[a]	GDM[b]	
Overall				
Laboratory	NGT	50	5	55
	GDM	2	4	6
Total	__[c]	52	9	61
kappa		0.457		
Fasting				
Laboratory	NGT	57	2	59
	GDM	1	3	4
Total	—	58	5	63
kappa		0.641		
2-hour				
Laboratory	NGT	52	4	56
	GDM	3	2	5
Total	—	55	6	61
kappa		0.301		

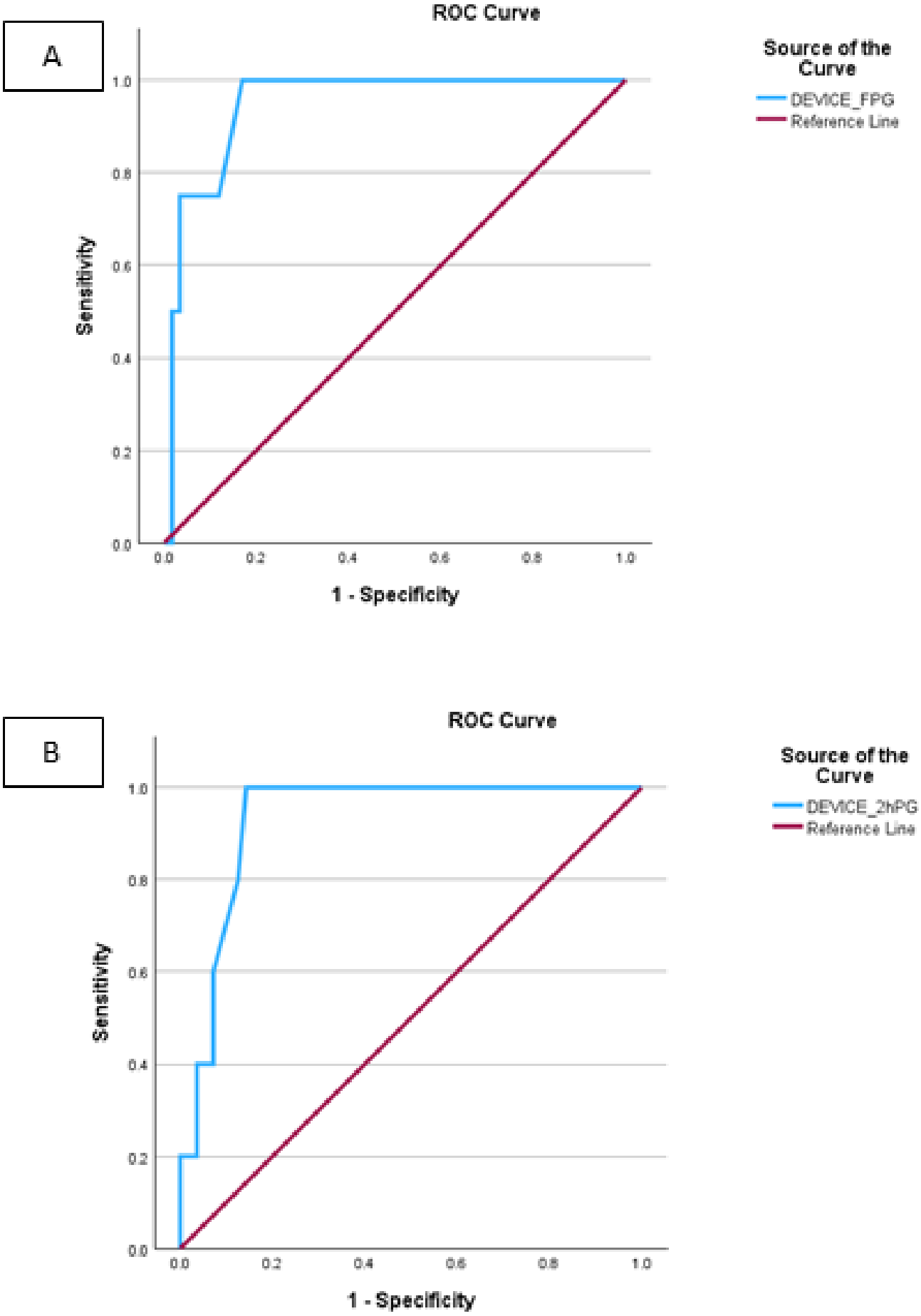
[a]NGT: normal glucose tolerance.

[b]GDM: gestational diabetes mellitus.

[c]Not applicable.

XSL•FO
RenderX

Figure 4. Receiver operating characteristic (ROC) curves with sensitivity/specificity. (A) Fasting glucose (cut-off 5.6 mmol/L); (B) 2-hour glucose (cut-off 7.8 mmol/L). FPG: fasting plasma glucose.



XSL•FO
RenderX

Table . Positive and negative predictive values.

	Positive predictive value	Negative predictive value
Overall	0.44	0.96
Fasting glucose	0.6	0.98
2-hour glucose	0.33	0.95

Discussion

Principal Findings and Comparison With Previous Works

In pregnant women in a clinical setting, the GTT@home OGTT device was compared with routine laboratory analysis of blood glucose samples. The GTT@home device worked well with relatively few failures. It showed a low bias (+0.16 mmol/L) across the range of all glucose concentrations, and for both fasting (+0.01 mmol/L) and 2-hour (+0.31 mmol/L) samples. Comparison of the classification of glucose intolerance between the GTT@home device and measurement using conventional laboratory glucose analysis showed small differences in the numbers of participants in the glucose intolerance categories.

These small differences may well reflect the tolerance around the different glucose analytical methods used. A previous study in which samples were analyzed at five central laboratories using four different automated glucose hexokinase methods demonstrated that despite there being low bias in glucose measurements across laboratories, the resulting GDM prevalence ranged considerably from 30.0% to 41.1% across laboratories [18]. Furthermore, even within the hospital clinical setting, the turn-around time of glucose samples can impact the clinical accuracy of laboratory measurements. Jangam et al [19] observed that delays of 15 min or more reduced clinical accuracy below 95%, and the accuracy was less than 65% for delays of 60 min. These processing delays in glucose measurements reduced the clinical relevance of results in patients with type 1 diabetes and were likely to similarly degrade the clinical value of measurements in other patient populations.

In community settings, pre-analytical issues and possible degradation of samples during transfer to the laboratory can severely affect the rates of diagnosis of GDM. Jamieson et al [20] showed an underdiagnosis rate of 62% due to the impact of long delays in centrifugation for OGTT samples in regional, rural, and remote sites in Western Australia. Potter et al [11] also showed that the variability in pre-analytic processing of blood for glucose measurement during pregnancy OGTTs could affect the GDM diagnostic rates, observing an increase in the rate of GDM from 11.6% to 20.6% by changing the process to centrifuging blood collected into sodium fluoride tubes, within 10 minutes of venipuncture.

In the context of our study, the relatively poor kappa statistics and positive predictive value score are likely to be due to the low number of participants with GDM included in the study cohort, small differences in glucose values close to the diagnostic cut-offs, and possibly due to the narrow range of glucose concentrations tested. In addition, pre-analytical issues and possible degradation of samples during transfer to the laboratory may also be responsible, as reflected by the device classifying slightly more individuals as having GDM than the laboratory method.

Conclusions

The GTT@home device worked well in a controlled, antenatal clinical setting. Differences in classification observed were likely due to pre-analytical issues associated with the laboratory tested samples. The GTT@home device therefore shows promise for home testing of glucose tolerance in pregnant women, in addition to wider community use. Further insight into the real-life usage of the device will be achieved with future studies by comparing an at-home diagnosis using the GTT@home device, performed by the OGTT recipient, within a few days of a routine in-hospital laboratory-measured OGTT.

Acknowledgments

We are grateful to all the participants who took part in the study and the staff at Neath Port Talbot Hospital for undertaking recruitment.

Funding

This clinical study was funded by Digostics Ltd.

Authors' Contributions

GJD, SDL, RP, and WYC designed the study. NJ, MC, MA, and RP were responsible for data collection. GJD, SDL, WYC, SNP, and RP were responsible for data analysis. All authors contributed to drafting, reviewing and approving the final version of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Centers for Disease Control and Prevention. National diabetes fact sheet: general information and national estimates on diabetes in the United States, 2007. : US Department of Health and Human Services, Centers for Disease Control and Prevention; 2008 URL: https://stacks.cdc.gov/view/cdc/5613/cdc_5613_DS1.pdf [accessed 2026-01-28]
2. NICE guideline NG3 diabetes in pregnancy: management from preconception to the postnatal period. NICE. 2015 Feb 25. URL: https://www.nice.org.uk/guidance/ng3 [accessed 2026-01-28]
3. van-de-l'Isle Y, Steer PJ, Watt Coote I, Cauldwell M. Impact of changes to national UK Guidance on testing for gestational diabetes screening during a pandemic: a single-centre observational study. BJOG 2021 Apr;128(5):917-920. [doi: 10.1111/1471-0528.16482] [Medline: 32888369]
4. Curtis AM, Farmer AJ, Roberts NW, Armitage LC. Performance of guidelines for the screening and diagnosis of gestational diabetes mellitus during the COVID-19 pandemic: a scoping review of the guidelines and diagnostic studies evaluating the recommended testing strategies. Diabet Epidemiol Manag 2021;3:100023. [doi: 10.1016/j.deman.2021.100023] [Medline: 35072134]
5. Bonetti G, Cancelli V, Coccoli G, et al. Which sample tube should be used for routine glucose determination? Prim Care Diabetes 2016 Jun;10(3):227-232. [doi: 10.1016/j.pcd.2015.11.003] [Medline: 26657574]
6. Norman M, Jones I. The shift from fluoride/oxalate to acid citrate/fluoride blood collection tubes for glucose testing - the impact upon patient results. Clin Biochem 2014 May;47(7-8):683-685. [doi: 10.1016/j.clinbiochem.2014.01.011] [Medline: 24463230]
7. Gambino R, Piscitelli J, Ackattupathil TA, et al. Acidification of blood is superior to sodium fluoride alone as an inhibitor of glycolysis. Clin Chem 2009 May;55(5):1019-1021. [doi: 10.1373/clinchem.2008.121707] [Medline: 19282354]
8. Lippi G, Nybo M, Cadamuro J, Guimaraes JT, van Dongen-Lases E, Simundic AM. Blood glucose determination: effect of tube additives. Adv Clin Chem 2018;84:101-123. [doi: 10.1016/bs.acc.2017.12.003] [Medline: 29478513]
9. van Balveren JA, Huijskens MJ, Gemen EF, Péquériaux NC, Kusters R. Effects of time and temperature on 48 routine chemistry, haematology and coagulation analytes in whole blood samples. Ann Clin Biochem 2017 Jul;54(4):448-462. [doi: 10.1177/0004563216665868] [Medline: 27497436]
10. Bruns DE, Metzger BE, Sacks DB. Diagnosis of gestational diabetes mellitus will be flawed until we can measure glucose. Clin Chem 2020 Feb 1;66(2):265-267. [doi: 10.1093/clinchem/hvz027] [Medline: 32040567]
11. Potter JM, Hickman PE, Oakman C, Woods C, Nolan CJ. Strict preanalytical oral glucose tolerance test blood sample handling is essential for diagnosing gestational diabetes mellitus. Diabetes Care 2020 Jul;43(7):1438-1441. [doi: 10.2337/dc20-0304] [Medline: 32350020]
12. Sacks DB, Arnold M, Bakris GL, et al. Position statement executive summary: guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Diabetes Care 2011 Jun;34(6):1419-1423. [doi: 10.2337/dc11-9997] [Medline: 21617111]
13. Sacks DB, Arnold M, Bakris GL, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2023 Aug 2;69(8):808-868. [doi: 10.1093/clinchem/hvad080] [Medline: 37473453]
14. Dunseath GJ, Bright D, Jones C, Dowrick S, Cheung WY, Luzio SD. Performance evaluation of a self‐administered home oral glucose tolerance test kit in a controlled clinical research setting. Diabet Med 2019 Jul;36(7):862-867. [doi: 10.1111/dme.13961] [Medline: 30972793]
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
16. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. J Diabetes Sci Technol 2014 Jul;8(4):658-672. [doi: 10.1177/1932296814539589] [Medline: 25562886]
17. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ 1994 Jul 9;309(6947):102. [doi: 10.1136/bmj.309.6947.102] [Medline: 8038641]
18. Scheuer CM, Tvarnø CD, Gils C, et al. The impact of inter-laboratory glucose bias on the diagnosis of gestational diabetes mellitus: comparison of common automated central laboratory methods. Clin Chim Acta 2023 Jun 1;546:117414. [doi: 10.1016/j.cca.2023.117414] [Medline: 37263535]
19. Jangam SR, Hayter G, Dunn TC. Impact of glucose measurement processing delays on clinical accuracy and relevance. J Diabetes Sci Technol 2013 May 1;7(3):660-668. [doi: 10.1177/193229681300700311] [Medline: 23759399]
20. Jamieson EL, Spry EP, Kirke AB, Atkinson DN, Marley JV. Real-world gestational diabetes screening: problems with the oral glucose tolerance test in rural and remote Australia. Int J Environ Res Public Health 2019 Nov 14;16(22):4488. [doi: 10.3390/ijerph16224488] [Medline: 31739513]

Abbreviations

GDM: gestational diabetes mellitus
NICE: National Institute for Health and Care Excellence
OGTT: oral glucose tolerance test

XSL•FO
RenderX

Edited by R Correa; submitted 05.Dec.2024; peer-reviewed by C Porter, SC Oglak; revised version received 30.Apr.2025; accepted 10.May.2025; published 11.Feb.2026.

Please cite as:
Dunseath GJ, Luzio SD, Cheung WY, Parsons SN, John N, Chokor M, Atkinson M, Peter R
Use of the GTT@home Oral Glucose Tolerance Test Kit in Gestational Diabetes Mellitus: Performance Evaluation Study
JMIR Diabetes 2026;11:e69695
URL: https://diabetes.jmir.org/2026/1/e69695
doi:10.2196/69695

© Gareth J Dunseath, Stephen D Luzio, Wai Yee Cheung, Sharon N Parsons, Nicola John, Mahmoud Chokor, Michael Atkinson, Rajesh Peter. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Social Determinants of Health Among Chinese Americans at Risk for Diabetes in a Mobile Diabetes Prevention Trial: Cross-Sectional Baseline Analysis

Nan Jiang[1], PhD; Jing Liu[1], PhD; Haili Song[1], MA; Yanan Zhao[1], PhD; Huilin Li[1], PhD; Stella S Yi[1], PMH, PhD; Jeannette M Beasley[2,3], RND, MPH, PhD; Lu Hu[1], PhD

[1]Department of Population Health, NYU Grossman School of Medicine, 180 Madison Avenue, New York, NY, United States

[2]Department of Medicine, NYU Grossman School of Medicine, New York, NY, United States

[3]Department of Nutrition and Food Studies, Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY, United States

Corresponding Author:
Lu Hu, PhD
Department of Population Health, NYU Grossman School of Medicine, 180 Madison Avenue, New York, NY, United States

Abstract

Background: Prediabetes is common in the United States, and adverse social determinants of health (SDOH) are known to undermine diabetes prevention efforts. Chinese Americans experience a disproportionately high prevalence of prediabetes, yet the SDOH profiles of this population remain understudied.

Objective: This study assessed SDOH among Chinese Americans at risk for diabetes and examined the associations between sociodemographic characteristics and SDOH barriers.

Methods: We conducted a cross-sectional analysis of baseline survey data from the Integrating Cultural Aspects into Diabetes Education (INCLUDE) study, a randomized controlled trial of a culturally and linguistically tailored mobile diabetes prevention intervention for Chinese Americans. Participants at risk for diabetes were enrolled between April 2023 and June 2024 in New York City (N=150). Measures included in the analyses were a 14-item SDOH scale (range 0-14, with higher scores indicating more barriers) and sociodemographic characteristics. Due to the small frequencies of high SDOH scores, we collapsed the outcome into 5 categories (0, 1, 2, 3, and 4-14) to improve model stability. We first used univariable logistic regression models to examine associations between each sociodemographic factor (age, sex, years of US residence, English proficiency, education, marital status, employment status, and annual household income) and the collapsed SDOH category, followed by a multivariable ordinal regression model including all sociodemographic variables.

Results: A total of 150 participants had a mean age of 49.9 (SD 12.6) years. Most were female (n=124, 82.7%), born outside the United States (n=149, 99.3%), and reported speaking English less than very well (n=132, 88.0%). Among respondents to the SDOH items (n=149), the mean SDOH score was 2.4 (SD 2.3), and 81.9% (n=122) reported at least 1 SDOH barrier. The three most frequently reported barriers were (1) the need to improve English proficiency, reading skills, or educational attainment (n=77, 51.7%); (2) experiences of racial discrimination (n=49, 32.9%); and (3) adverse housing conditions (n=38, 25.5%). After collapsing the original SDOH score, 27 (18.2%) participants had a score of 0, 39 (26.2%) had a score of 1, 23 (15.4%) had a score of 2, 27 (18.2%) had a score of 3, and 33 (22.1%) had scores of 4 to 14. In the multivariable analysis, female sex (vs male) was associated with higher SDOH score categories (odds ratio 3.83, 95% CI 1.65-9.16; P=.002).

Conclusions: SDOH-related barriers were prevalent among Chinese Americans at risk for diabetes. Diabetes prevention efforts should incorporate routine SDOH screening and structured resource navigation or referral pathways, with particular attention to subgroups at higher risk, such as female individuals.

Trial Registration: ClinicalTrials.gov NCT05492916; https://clinicaltrials.gov/study/NCT05492916

International Registered Report Identifier (IRRID): RR2-10.2196/65455

(JMIR Diabetes 2026;11:e95295) doi:10.2196/95295

KEYWORDS

prediabetes; diabetes; social determinants of health; Chinese American; mobile health

XSL•FO
RenderX

Introduction

Prediabetes imposes a substantial and growing health and economic burden in the United States [1]. National medical costs attributable to prediabetes were estimated at US $25 billion in 2007 and US $44 billion in 2012, representing a 74% increase over 5 years [2,3]. Prediabetes is a state of impaired glucose regulation, with blood glucose levels above normal but below diagnostic thresholds for diabetes [4]. In 2023, an estimated 115.2 million US adults aged ≥18 years (43.5% of the adult population) had prediabetes [5], and approximately 5% to 10% of individuals with prediabetes progress to type 2 diabetes annually [4,6]. Therefore, preventive strategies targeting individuals with prediabetes are essential to reduce progression to diabetes.

Social determinants of health (SDOH) refer to the conditions in which people are born, grow, live, work, and age [7]. Healthy People 2030 organizes SDOH into five domains: (1) economic stability, (2) education access and quality, (3) health care access and quality, (4) neighborhood and built environment, and (5) social and community context [8]. Adverse SDOH (eg, housing instability, food insecurity, limited insurance coverage, and restricted access to health care) affect diabetes outcomes by posing barriers to effective self-management, including healthy diet, physical activity, medication adherence, and glucose monitoring [9-12]. Evidence shows that adverse SDOH are associated with an increased risk of diabetes [13-15].

Chinese Americans represent the largest Asian American ethnic group in the United States, with approximately 5.5 million individuals, including 2.9 million who were born outside the United States [16,17]. Chinese Americans also experience a disproportionately high prevalence of prediabetes. The burden of diabetes and prediabetes is substantial in this population. Among 373,098 adults aged 45 to 64 years in the Kaiser Permanente Northern California health plan, the prevalence of prediabetes was significantly higher among Chinese Americans than among White individuals (37.8% vs 24.5% among men; 30.8% vs 18.0% among women) [18]. Chinese Americans, particularly immigrants, often face SDOH barriers, including lower education and income, limited English proficiency, and a lack of health insurance [19-21]. These barriers may impede access to diabetes prevention programs and increase the risk of progression from prediabetes to diabetes.

Despite growing attention to SDOH in diabetes prevention and care, little is known about SDOH profiles among Chinese Americans at risk for diabetes. To address this gap, we conducted an exploratory cross-sectional analysis of baseline data from a randomized controlled trial of a culturally and linguistically tailored, digital diabetes prevention intervention for Chinese Americans at risk for diabetes. The objectives of this study were to describe SDOH status and assess associations between sociodemographic characteristics and SDOH barriers within this sample. The findings will provide insights into SDOH patterns in an understudied population and may inform future diabetes prevention efforts among Chinese Americans.

Methods

Study Design

We analyzed baseline survey data from the Integrating Cultural Aspects into Diabetes Education (INCLUDE) trial, a randomized controlled trial evaluating the efficacy of a culturally and linguistically tailored mobile diabetes prevention program for Chinese Americans at risk of diabetes, with weight loss as the primary outcome. Details about the INCLUDE trial have been reported elsewhere [22].

Ethical Considerations

The study protocol was approved by the New York University Grossman School of Medicine Institutional Review Board (s22-00783) and was registered at ClinicalTrials.gov (NCT05492916). All participants were fully informed, both verbally and in writing, of the study purpose, procedures, potential risks and benefits, the voluntary nature of participation, and their right to withdraw. Participant privacy and confidentiality were strictly maintained throughout the study. Written informed consent was obtained from all participants prior to the baseline survey. Participants received a US $25 gift card as compensation for completing the baseline survey.

Participants and Recruitment

A total of 150 participants residing in New York City (NYC) were enrolled in the INCLUDE trial between April 2023 and June 2024. Eligibility criteria were as follows: (1) self-identification as being of Chinese ethnicity; (2) aged 18 to 70 years; (3) diagnosis of prediabetes or a score of ≥5 on a validated prediabetes risk test [23]; (4) BMI ≥23 kg/m², using the Asian-specific overweight threshold [24]; (5) willingness to receive program videos via SMS text messaging or WeChat (Tencent Holdings Ltd); and (6) smartphone ownership or willingness to use a study-provided smartphone. Exclusion criteria were as follows: (1) inability or unwillingness to provide informed consent, (2) hearing or vision impairment that precluded participation, (3) refusal of randomization, and (4) current pregnancy or breastfeeding or plans to become pregnant within the next 6 months.

Participants were recruited using five primary strategies: (1) partnerships with community-based organizations and health care facilities serving Chinese American communities in NYC to disseminate study flyers (eg, posting flyers in offices, on organizational websites, and through WeChat Moments); (2) advertisements in local newspapers and other media outlets; (3) community outreach events (eg, food pantries, health fairs, and health workshops); (4) proactive outreach using the New York University Langone Health Epic electronic health record across Manhattan, Queens, Brooklyn, and Long Island (potentially eligible individuals were identified, mailed an invitation letter describing the study, and then contacted by phone); and (5) peer referrals, in which individuals voluntarily shared study flyers within their social networks without financial incentives.

Screening, Enrollment, and Baseline Survey

Interested individuals contacted bilingual research assistants (RAs) by calling the study phone number or approaching RAs

XSL•FO
RenderX

at in-person recruitment activities (eg, community outreach events). Trained RAs conducted eligibility screening by phone or in person. For eligible participants, RAs obtained informed consent (verbal consent for phone-based sessions and written consent for in-person sessions) and administered the baseline survey by phone or in person. Participants could complete the consent process and the survey in Chinese or English. Participants received the compensation after completing the baseline survey.

Measures

The analysis included a 14-item SDOH measure and sociodemographic characteristics. The 14-item SDOH instrument was adapted from validated tools, including the brief food insecurity screener [25] and domains from the Accountable Health Communities Health-Related Social Needs Screening Tool [26]. This instrument underwent a standard translation and back-translation process to ensure linguistic accuracy and cultural appropriateness for this population. It assessed core health-related barriers and social needs across 5 domains, including economic stability (eg, food access and housing stability), education access and quality (eg, need to improve English proficiency, reading skills, or educational attainment), health care access and quality (eg, insurance coverage), neighborhood and built environment (eg, housing conditions), and social and community context (eg, racial discrimination; need for childcare; and need for help addressing tobacco, alcohol, or drug use). Items had dichotomous response options (1="Yes" and 0="No"). An overall SDOH score was calculated by summing the responses across the 14 items (range 0-14), with higher scores indicating more SDOH barriers. Given the relatively small sample size and the exploratory nature of the analysis, we used a composite score to capture the overall burden of SDOH barriers rather than performing domain-specific analyses to maintain statistical power.

Sociodemographic characteristics included age, sex (female or male), country of birth (born outside the United States vs born in the United States), years of residence in the United States, self-reported English proficiency ("very well," "well," "not well," and "not at all"), education (less than high school, high school graduate, and more than high school), marital status (married or cohabitating vs other), employment status (employed, unemployed, and retired), annual household income (US <$25,000, US $25,000-$55,000, US >$55,000, and "unreported/don't know"), and health insurance type (Medicaid, Medicare, private insurance, other public or government insurance, employer-sponsored insurance, and uninsured). For health insurance type, participants could select all applicable responses.

Statistical Analysis

Descriptive statistics summarized sociodemographic characteristics and the SDOH score, with categorical variables reported as frequencies and percentages and continuous variables reported as means (SDs). Before examining associations between sociodemographic factors and the SDOH score, we first assessed the distribution of SDOH scores among respondents who completed the SDOH measures (n=149). The score ranged from 0 to 14 (15 possible values), although only 11 outcome scores were observed, with no participants reporting scores of 9, 11, 13, or 14. Table S1 in Multimedia Appendix 1 presents these observed scores and participant characteristics for each score. Given the relatively small sample size (n=149) and the number of observed outcome scores, the proportional odds assumption could be restrictive, and model estimates might be unstable because of sparse data for some scores. Guided by the score distribution shown in Table S1 in Multimedia Appendix 1, we collapsed the 11 observed outcome scores into 5 categories (0, 1, 2, 3, and 4-14) to improve model stability while preserving meaningful variability in the outcome.

The collapsed SDOH category was modeled using proportional odds logistic regression, which estimates the log-odds of being in a higher category of the ordinal outcome relative to all lower categories cumulatively. This model assumes that the association between each demographic variable and the collapsed SDOH category is constant across all outcome thresholds (the proportional odds assumption). Covariates included age, sex, years of US residence, English proficiency, education, marital status, employment status, and annual household income. Health insurance type was not included in the regression analyses because participants could select multiple categories. The model was specified as follows:

$$\log (\Pr(Y\leq k)\Pr(Y>k))=\alpha k+\beta TX,\ k=1,\ldots,K-1,$$

where Y denotes the collapsed SDOH category, and X denotes covariates of interest. β represents regression coefficients, and α_k represents threshold-specific intercepts. We first fitted univariable models to examine associations between each sociodemographic factor and the collapsed SDOH category. We then fitted a multivariable model including all sociodemographic factors to assess their independent associations with the collapsed SDOH category. Before conducting the analyses, the proportional odds assumption was evaluated for all ordinal regression models using the Brant test implemented in R (version 4.5.1; R Foundation for Statistical Computing), confirming that the assumption of parallel regression coefficients was reasonable for the included variables (Table S2 in Multimedia Appendix 1). The proportional odds assumption was evaluated for all ordinal regression models using the nominal effects test, and no evidence of violation was detected. Odds ratios (ORs) were obtained by exponentiating the estimated regression coefficients (exp [β]). Corresponding 95% CIs were calculated as exp($\hat{\beta}\pm1.96\times SE$), and Wald tests were used to assess statistical significance. All analyses were conducted using R software with a 2-sided significance level of .05.

We also conducted univariable and multivariable regression analyses using the original SDOH score as the outcome to examine associations between sociodemographic characteristics and SDOH. Incidence rate ratios, corresponding 95% CIs, and P values were calculated. We conducted sensitivity analyses using a negative binomial model to evaluate the robustness of our results. The negative binomial model was used because the SDOH score is a count variable that may exhibit overdispersion, making this model appropriate for such outcomes. For univariable and multivariable analyses, we also evaluated model fit using Akaike Information Criterion (AIC) and Bayesian

XSL•FO
RenderX

Information Criterion (BIC) for both the proportional odds regression model with the collapsed SDOH categories and the negative binomial regression model with the original SDOH score. Because of the small cell size for the "unreported/don't know" household income category (10/149, 6.7%), we conducted sensitivity analyses excluding these participants and treating this response as missing rather than as a separate category. We then reran the multivariable analyses using (1) proportional odds regression with collapsed SDOH categories and (2) negative binomial regression with the original SDOH score.

Results

Participant Characteristics

A total of 150 participants had a mean age of 49.9 (SD 12.6) years (Table 1). Most were female (n=124, 82.7%) and born outside the United States (n=149, 99.3%), including 129 (86.0%) from mainland China and 15 (10.0%) from Hong Kong (data not shown). The mean length of US residence was 19.4 (SD 10.6) years. Overall, 132 (88.0%) participants reported speaking English less than very well, 81 (54.0%) had a high school education or less, 116 (77.3%) were married or cohabitating, 102 (68.0%) were employed, 56 (37.3%) reported an annual household income of US <$25,000, and 75 (50.0%) were insured through Medicaid.

XSL•FO
RenderX

Table . Participant characteristics of Chinese Americans at risk for diabetes (N=150).

Characteristics	Values
Age (years), mean (SD)	49.9 (12.6)
Female sex, n (%)	124 (82.7)
Born outside the United States, n (%)	149 (99.3)
Period of residence in the United States (years), mean (SD)	19.4 (10.6)
English proficiency, n (%)	
Very well	18 (12.0)
Well	50 (33.3)
Not well	68 (45.3)
Not at all	14 (9.3)
Education, n (%)	
Less than high school	41 (27.3)
High school graduate	40 (26.7)
More than high school	69 (46.0)
Marital status, n (%)	
Married or cohabitating	116 (77.3)
Other	34 (22.7)
Employment status, n (%)	
Employed	102 (68.0)
Unemployed	20 (13.3)
Retired	28 (18.7)
Annual household income (US $), n (%)	
<25,000	56 (37.3)
25,000-55,000	47 (31.3)
>55,000	37 (24.7)
Unreported/don't know	10 (6.7)
Insurance type[a], n (%)	
Medicaid	75 (50.0)
Medicare	18 (12.0)
Private insurance	20 (13.3)
Other insurance (public or government)	26 (17.3)
Employee-sponsored insurance	22 (14.7)
Uninsured	6 (4.0)

[a]Participants could select multiple responses; therefore, percentages do not total 100%.

SDOH Measures Analysis

Among the 149 respondents who completed the SDOH measures, the mean SDOH score was 2.4 (SD 2.3), and 81.9% (n=122) of participants reported at least 1 SDOH barrier (Table 2). The five most frequently reported barriers were (1) the need to improve English proficiency, reading skills, or educational attainment (n=77, 51.7%); (2) experiences of racial discrimination (n=49, 32.9%); (3) adverse housing conditions (eg, mold, pests, or structural damage; n=38, 25.5%); (4) the need for welfare support (eg, health insurance; n=26, 17.4%); and (5) concern about child behavioral problems (n=26, 17.4%). After collapsing the original SDOH scores into 5 categories, 27 (18.2%) participants had a score of 0, 39 (26.2%) had a score of 1, 23 (15.4%) had a score of 2, 27 (18.2%) had a score of 3, and 33 (22.1%) had scores of 4 to 14 (Table 3).

XSL•FO
RenderX

Table . Social determinants of health (SDOH) among respondents who completed the SDOH measures (N=149).

SDOH measures	Values
Overall SDOH score, mean (SD)	2.4 (2.3)
Participants reporting any SDOH-related barriers, n (%)	122 (81.9)
Need to improve English proficiency, reading skills, or educational attainment	77 (51.7)
Experiences of racial discrimination	49 (32.9)
Adverse housing conditions (eg, mold, pests, or structural damage)	38 (25.5)
Need for welfare support (eg, benefits enrollment or health insurance)	26 (17.4)
Concern about child behavioral problems (eg, tantrums or hitting)	26 (17.4)
Need for childcare support	23 (15.4)
Housing insecurity or concern about eviction or utility shutoff	23 (15.4)
Food insufficiency (food running out due to lack of money)	23 (15.4)
Transportation barriers affecting appointments, work, or essential activities	22 (14.8)
Worry about running out of food due to financial stress	20 (13.4)
Need for essential child supplies (eg, diapers, car seat, crib, or stroller)	14 (9.4)
Need for legal assistance (eg, immigration or custody)	11 (7.4)
Need for help addressing tobacco, alcohol, or drug use	9 (6.0)
Domestic violence–related safety concerns at home	2 (1.3)

XSL•FO
RenderX

Table . Sociodemographic characteristics by collapsed social determinants of health (SDOH) categories among respondents who completed the SDOH measures (N=149).

Characteristics	Score 0 (n=27)	Score 1 (n=39)	Score 2 (n=23)	Score 3 (n=27)	Score 4-14 (n=33)
Age (years), mean (SD)	52.3 (11.1)	51.3 (13.3)	50.7 (11.9)	49.4 (14.3)	46.8 (12.2)
Sex, n (%)					
Female	19 (70.4)	27 (69.2)	21 (91.3)	24 (88.9)	32 (97.0)
Male	8 (29.6)	12 (30.8)	2 (8.7)	3 (11.1)	1 (3.0)
Period of US residence (years), mean (SD)	24.0 (9.2)	19.6 (10.4)	17.5 (12.1)	22.7 (11.7)	14.2 (7.5)
English proficiency, n (%)					
Very well	6 (22.2)	6 (15.4)	1 (4.3)	5 (18.5)	0 (0)
Well	11 (40.7)	12 (30.8)	6 (26.1)	9 (33.3)	12 (36.4)
Not well	8 (29.6)	16 (41.0)	13 (56.5)	12 (44.4)	18 (54.5)
Not at all	2 (7.4)	5 (12.8)	3 (13.0)	1 (3.7)	3 (9.1)
Education, n (%)					
Less than high school	4 (14.8)	12 (30.8)	11 (47.8)	5 (18.5)	9 (27.3)
High school graduate	7 (25.9)	8 (20.5)	3 (13.0)	9 (33.3)	13 (39.4)
More than high school	16 (59.3)	19 (48.7)	9 (39.1)	13 (48.1)	11 (33.3)
Marital status, n (%)					
Married or cohabitating	16 (59.3)	37 (94.9)	16 (69.6)	19 (70.4)	27 (81.8)
Other	11 (40.7)	2 (5.1)	7 (30.4)	8 (29.6)	6 (18.2)
Employment status, n (%)					
Employed	20 (74.1)	27 (69.2)	14 (60.9)	19 (70.4)	21 (63.6)
Unemployed	2 (7.4)	3 (7.7)	4 (17.4)	2 (7.4)	9 (27.3)
Retired	5 (18.5)	9 (23.1)	5 (21.7)	6 (22.2)	3 (9.1)
Annual household income (US $), n (%)					
<25,000	10 (37.0)	14 (35.9)	10 (43.5)	9 (33.3)	13 (39.4)
25,000-55,000	5 (18.5)	10 (25.6)	9 (39.1)	7 (25.9)	15 (45.5)
>55,000	11 (40.7)	10 (25.6)	4 (17.4)	8 (29.6)	4 (12.1)
Unreported/don't know	1 (3.7)	5 (12.8)	0 (0)	3 (11.1)	1 (3.0)

Associations Between Sociodemographic Characteristics and SDOH

In univariable regression analyses (Table 4), female sex (vs male; OR 3.77, 95% CI 1.77-8.20; *P*=.001), self-reported speaking English "not well" (vs "very well"; OR 3.35, 95% CI 1.33-8.69; *P*=.01), and being unemployed (vs employed; OR 2.62, 95% CI 1.08-6.58; *P*=.04) were associated with higher SDOH score categories (indicating more barriers), while longer residence in the United States (OR 0.96, 95% CI 0.94-0.99; *P*=.007) was associated with lower SDOH score categories. In multivariable analyses, female sex (OR 3.83, 95% CI 1.65-9.16; *P*=.002) remained independently associated with higher SDOH score categories.

XSL•FO
RenderX

Table . Associations between sociodemographic characteristics and collapsed social determinants of health categories.

Characteristics	Univariable analysis		Multivariable analysis	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)	0.98 (0.96-1.01)	.08	0.97 (0.94-1.01)	.08
Female (reference: male)	3.77[b] (1.77-8.20)	.001	3.83[b] (1.65-9.16)	.002
Period of residence in the United States (years)	0.96[b] (0.94-0.99)	.007	0.97 (0.94-1.01)	.14
English proficiency (reference: very well)				
Well	2.39 (0.91-6.41)	.08	1.68 (0.56-5.17)	.36
Not well	3.35[c] (1.33-8.69)	.01	2.38 (0.62-9.26)	.21
Not at all	2.13 (0.63-7.39)	.23	2.69 (0.45-16.14)	.28
Education (reference: less than high school)				
High school graduate	1.38 (0.63-2.99)	.41	1.50 (0.63-3.59)	.36
More than high school	0.68 (0.34-1.30)	.25	1.07 (0.41-2.82)	.90
Marital status: other (reference: married or cohabitating)	0.80 (0.39-1.60)	.52	0.76 (0.35-1.64)	.49
Employment status (reference: employed)				
Unemployed	2.62[c] (1.08-6.58)	.04	2.27 (0.89-6.07)	.09
Retired	0.82 (0.40-1.68)	.58	1.60 (0.60-4.29)	.34
Annual household income (US $; reference: <25,000)				
25,000-55,000	1.57 (0.78-3.16)	.21	1.27 (0.53-3.09)	.59
>55,000	0.55 (0.26-1.16)	.12	0.78 (0.25-2.41)	.67
Unreported/don't know	0.77 (0.24-2.45)	.66	0.91 (0.22-3.74)	.90

[a]OR: odds ratio.

[b]*P*<.01.

[c]*P*<.05.

Sensitivity Analyses

Using the original SDOH score as the outcome, Table S3 in Multimedia Appendix 1 shows that the associations between demographic variables and SDOH were generally consistent in direction and magnitude with those observed in the primary analyses. In the multivariable negative binomial model, female participants had significantly higher SDOH scores than male participants (incidence rate ratio 2.33, 95% CI 1.46-3.70; *P*<.001).

Table S4 in Multimedia Appendix 1 presents the AIC and BIC values for the proportional odds logistic regression models with collapsed SDOH categories and the negative binomial regression models with the original SDOH score. Across both univariable and multivariable analyses, the proportional odds models with collapsed SDOH categories generally showed lower AIC and BIC values than the negative binomial models, suggesting better model fit. For example, in the multivariable analysis, the AIC and BIC for the proportional odds model were 475.80 and 529.87, respectively, compared with 595.58 and 643.65 for the negative binomial model. These findings suggest that the proportional odds models provided a more parsimonious fit to the data while adequately capturing variation in SDOH outcomes.

Table S5 in Multimedia Appendix 1 shows that, when the "unreported/don't know" household income group (10/149, 6.7%) was excluded from the multivariable analyses, the findings remained similar to those from the primary analyses. Specifically, the direction, magnitude, and statistical interpretation of the associations between sociodemographic factors and both the collapsed SDOH categories and the original SDOH score were largely unchanged. These results further support the robustness of the study findings.

XSL•FO
RenderX

Discussion

Principal Findings

This study examined SDOH in a NYC community-based sample of Chinese Americans at risk for diabetes. It is among the first studies to assess SDOH in this understudied population in diabetes research [27]. The "model minority" stereotype often portrays Chinese Americans as uniformly advantaged, self-sufficient, and living under relatively favorable socioeconomic and health circumstances [28]. Our findings challenge this narrative by revealing significant socioeconomic disadvantage and multiple SDOH barriers in this sample, underscoring that Chinese Americans at risk for diabetes are not a uniformly advantaged group.

Most participants (122/149, 81.9%) reported at least 1 SDOH barrier. This finding is consistent with that of a prior study of safety-net patients with uncontrolled diabetes, in which 92% of participants reported at least 1 SDOH-related barrier [29]. The most commonly reported barriers included needs related to improving English proficiency, reading skills, or educational attainment (77/149, 51.7%), experiences of racial discrimination (49/149, 32.9%), adverse housing conditions (38/149, 25.5%), the need for welfare support (26/149, 17.4%), and concerns about child behavioral problems (26/149, 17.4%). More than one-fifth of participants (33/149, 22.1%) reported 4 or more SDOH barriers. These findings highlight the multiple social and structural challenges faced by this group and suggest that diabetes prevention programs may benefit from routine SDOH screening combined with structured resource navigation and referral pathways, such as access to health-related social services. These results further challenge prevailing "model minority" narratives and underscore the heterogeneity of needs within this population [21,30]. Without data of this type, the challenges faced by many Chinese Americans may remain obscured, potentially contributing to the underallocation of resources and an increased risk of progression from prediabetes to diabetes [31].

We also found that female sex was associated with more SDOH barriers. One possible explanation is that women may experience greater barriers because of caregiving responsibilities and related constraints on employment [32] and access to other resources. However, the male reference group in this study was small (n=26), which may have reduced the precision of comparisons by sex. Accordingly, the mechanisms underlying the higher SDOH scores observed among female participants should be interpreted cautiously and warrant further investigation in larger samples. Therefore, diabetes prevention programs serving Chinese Americans may benefit from targeted support for subgroups facing higher risk, such as women.

The associations of English proficiency, length of US residence, and employment status with SDOH barriers observed in univariable analyses were attenuated in the multivariable model. This may reflect confounding and shared variance with other sociodemographic variables. Further research in larger samples is warranted to examine their independent associations with SDOH among Chinese Americans at risk for diabetes.

Limitations

This study has several limitations. First, as an exploratory analysis of data from a single clinical trial, participants were recruited in NYC and were predominantly born outside the United States, which may limit generalizability to US-born Chinese Americans and to individuals living in other geographic regions. However, this concern is partly mitigated by the relatively long mean duration of US residence (19.4, SD 10.6 years) among participants born outside the United States. Second, the cross-sectional nature of the baseline data precludes causal inference, and self-reported measures are subject to recall and reporting bias. Finally, the modest sample size precluded domain-specific analyses of SDOH. Despite these limitations, this study provides a comprehensive assessment of SDOH in an understudied population at elevated risk for diabetes. Additionally, community-based recruitment complements health care–based samples and provides community-grounded insights to inform tailored strategies for diabetes prevention and care for Chinese Americans.

Conclusions

SDOH-related barriers are prevalent among Chinese Americans at risk for diabetes. Findings from this NYC community–based sample have important implications for diabetes prevention and care. Programs serving this population should not rely on "model minority" assumptions and may benefit from integrating routine SDOH screening with structured resource navigation and referral pathways to address identified needs. Targeted supports may be particularly important for subgroups facing greater risk, including female individuals. Future studies in larger Chinese American populations are needed to confirm these findings.

Acknowledgments

ChatGPT (version 5.5; OpenAI) was used to assist with language editing and manuscript refinement. All content was reviewed and approved by the authors.

Funding

This work was supported by the American Diabetes Association (grant 7-22-ICTSN-08). LH also received support from the National Institutes of Health (grants R01MD017579, K99MD012811, R00MD012811, P50MD017356 pilot award, U54MD000538-15 pilot award, and P30DK111022 pilot award) and the Agency for Healthcare Research and Quality (grant R18HS029813). The funding organizations had no role in the study design, data collection and analysis, manuscript writing, or the decision to submit the paper for publication.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: LH
Data curation: JL, YZ, HL, LH
Formal analysis: YZ, HL
Funding acquisition: LH
Investigation: JL, HS, YZ, HL, LH
Methodology: NJ, JL, HS, YZ, HL, LH
Project administration: LH
Resources: JL, YZ, HL, LH
Software: YZ, HL
Supervision: LH
Validation: YZ, HL, LH
Visualization: NJ
Writing—original draft: NJ, HS, YZ
Writing—review and editing: NJ, JL, HS, YZ, HL, SSY, JMB, LH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 59 KB - diabetes_v11i1e95295_app1.docx]

References

1. Cefalu WT, Petersen MP, Ratner RE. The alarming and rising costs of diabetes and prediabetes: a call for action!. Diabetes Care 2014 Dec;37(12):3137-3138. [doi: 10.2337/dc14-2329] [Medline: 25414386]
2. Dall TM, Zhang Y, Chen YJ, Quick WW, Yang WG, Fogli J. The economic burden of diabetes. Health Aff (Millwood) 2010 Feb;29(2):297-303. [doi: 10.1377/hlthaff.2009.0155] [Medline: 20075080]
3. Dall TM, Yang W, Halder P, et al. The economic burden of elevated blood glucose levels in 2012: diagnosed and undiagnosed diabetes, gestational diabetes mellitus, and prediabetes. Diabetes Care 2014 Dec;37(12):3172-3179. [doi: 10.2337/dc14-1036] [Medline: 25414388]
4. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet 2012 Jun 16;379(9833):2279-2290. [doi: 10.1016/S0140-6736(12)60283-9] [Medline: 22683128]
5. National diabetes statistics report. : Centers for Disease Control and Prevention URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2026-05-28]
6. Nathan DM, Davidson MB, DeFronzo RA, et al. Impaired fasting glucose and impaired glucose tolerance: implications for care. Diabetes Care 2007 Mar;30(3):753-759. [doi: 10.2337/dc07-9920] [Medline: 17327355]
7. Social determinants of health. World Health Organization. URL: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1 [accessed 2025-10-18]
8. Social determinants of health. U.S. Department of Health and Human Services. URL: https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health#:~:text=Healthy%20People%202030%20has%20an%20increased%20focus,built%20environment%20*%20Social%20and%20community%20context [accessed 2025-10-18]
9. Frier A, Devine S, Barnett F, McBain-Rigg K, Dunning T. Improving type 2 diabetes care and self-management at the individual level by incorporating social determinants of health. Aust N Z J Public Health 2022 Dec;46(6):865-871. [doi: 10.1111/1753-6405.13296] [Medline: 36192823]
10. Woodward A, Walters K, Davies N, et al. Barriers and facilitators of self-management of diabetes amongst people experiencing socioeconomic deprivation: a systematic review and qualitative synthesis. Health Expect 2024 Jun;27(3):e14070. [doi: 10.1111/hex.14070] [Medline: 38751247]
11. Hu J, Amirehsani K, Wallace DC, Letvak S. Perceptions of barriers in managing diabetes: perspectives of Hispanic immigrant patients and family members. Diabetes Educ 2013;39(4):494-503. [doi: 10.1177/0145721713486200] [Medline: 23640301]
12. Martinez-Cardoso A, Jang W, Baig AA. Moving diabetes upstream: the social determinants of diabetes management and control among immigrants in the US. Curr Diab Rep 2020 Aug 28;20(10):48. [doi: 10.1007/s11892-020-01332-w] [Medline: 32857197]

XSL•FO
RenderX

13. Hacker K, Thomas CW, Zhao G, Claxton JS, Eke P, Town M. Social determinants of health and health-related social needs among adults with chronic diseases in the United States, Behavioral Risk Factor Surveillance System, 2022. Prev Chronic Dis 2024 Nov 27;21:E94. [doi: 10.5888/pcd21.240362] [Medline: 39602222]
14. McClintock HF, Edmonds SE, Wang E. Racial/ethnic disparities in access to transportation among persons with type 2 diabetes mellitus. J Racial Ethn Health Disparities 2026 Feb;13(1):494-504. [doi: 10.1007/s40615-024-02262-z] [Medline: 39804504]
15. Hill-Briggs F, Fitzpatrick SL. Overview of social determinants of health in the development of diabetes. Diabetes Care 2023 Sep 1;46(9):1590-1598. [doi: 10.2337/dci23-0001] [Medline: 37354331]
16. 2023 ACS 1-year estimates. United States Census Bureau. 2023. URL: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2023/1-year.html [accessed 2026-5-28]
17. Facts about Chinese in the U.S. Pew Research Center. 2025. URL: https://www.pewresearch.org/race-and-ethnicity/fact-sheet/asian-americans-chinese-in-the-u-s/ [accessed 2025-10-18]
18. Vicks WS, Lo JC, Guo L, et al. Prevalence of prediabetes and diabetes vary by ethnicity among U.S. Asian adults at healthy weight, overweight, and obesity ranges: an electronic health record study. BMC Public Health 2022 Oct 22;22(1):1954. [doi: 10.1186/s12889-022-14362-8] [Medline: 36273116]
19. Greene M, Batalova J. Chinese immigrants in the United States. Migration Policy Institute. 2025. URL: https://www.migrationpolicy.org/article/chinese-immigrants-united-states [accessed 2026-05-28]
20. Wang A. Chinese in NYC: a profile. Asian American Federation. 2022. URL: https://www.aafederation.org/research/chinese-in-nyc-a-profile/ [accessed 2023-04-26]
21. Tan C, Wyatt LC, Kranick JA, Kwon SC, Oyebode O. Factors associated with health insurance status in an Asian American population in New York City: analysis of a community-based survey. J Racial Ethn Health Disparities 2018 Dec;5(6):1354-1364. [doi: 10.1007/s40615-018-0485-y] [Medline: 29582383]
22. Hu L, Lin NF, Shi Y, et al. The Integrating Cultural Aspects Into Diabetes Education (INCLUDE) study to prevent diabetes in Chinese immigrants: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Nov 19;13:e65455. [doi: 10.2196/65455] [Medline: 39560984]
23. Bang H, Edwards AM, Bomback AS, et al. Development and validation of a patient self-assessment score for diabetes risk. Ann Intern Med 2009 Dec 1;151(11):775-783. [doi: 10.7326/0003-4819-151-11-200912010-00005] [Medline: 19949143]
24. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
25. Hager ER, Quigg AM, Black MM, et al. Development and validity of a 2-item screen to identify families at risk for food insecurity. Pediatrics 2010 Jul;126(1):e26-e32. [doi: 10.1542/peds.2009-3146] [Medline: 20595453]
26. The Accountable Health Communities Health-Related Social Needs Screening Tool. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/priorities/innovation/files/worksheets/ahcm-screeningtool.pdf [accessed 2026-03-12]
27. Rajpathak SN, Wylie-Rosett J. High prevalence of diabetes and impaired fasting glucose among Chinese immigrants in New York City. J Immigr Minor Health 2011 Feb;13(1):181-183. [doi: 10.1007/s10903-010-9356-2] [Medline: 20533090]
28. Yi SS, Kwon SC, Sacks R, Trinh-Shevrin C. Commentary: persistence and health-related consequences of the model minority stereotype for Asian Americans. Ethn Dis 2016 Jan 21;26(1):133-138. [doi: 10.18865/ed.26.1.133] [Medline: 26843806]
29. Levy NK, Park A, Solis D, et al. Social determinants of health and diabetes-related distress in patients with insulin-dependent type 2 diabetes: cross-sectional, mixed methods approach. JMIR Form Res 2022 Oct 12;6(10):e40164. [doi: 10.2196/40164] [Medline: 36222807]
30. Chesla CA, Chun KM, Kwan CM, et al. Testing the efficacy of culturally adapted coping skills training for Chinese American immigrants with type 2 diabetes using community-based participatory research. Res Nurs Health 2013 Aug;36(4):359-372. [doi: 10.1002/nur.21543] [Medline: 23606271]
31. Min LY, Islam RB, Gandrakota N, Shah MK. The social determinants of health associated with cardiometabolic diseases among Asian American subgroups: a systematic review. BMC Health Serv Res 2022 Feb 25;22(1):257. [doi: 10.1186/s12913-022-07646-7] [Medline: 35216607]
32. Pendergrast C. Women report worse employment impacts from family caregiving. Maxwell School of Citizenship and Public Affairs, Syracuse University. 2021. URL: https://www.maxwell.syr.edu/research/lerner-center/population-health-research-brief-series/article/women-report-worse-employment-impacts-from-family-caregiving [accessed 2026-04-08]

Abbreviations

AIC: Akaike Information Criterion
BIC: Bayesian Information Criterion
INCLUDE: Integrating Cultural Aspects into Diabetes Education
NYC: New York City
OR: odds ratio

XSL•FO
RenderX

RA: research assistant
SDOH: social determinants of health

Edited by S Li; submitted 13.Mar.2026; peer-reviewed by H Wu, Q Fan; revised version received 07.May.2026; accepted 10.May.2026; published 03.Jun.2026.

Please cite as:
Jiang N, Liu J, Song H, Zhao Y, Li H, Yi SS, Beasley JM, Hu L
Social Determinants of Health Among Chinese Americans at Risk for Diabetes in a Mobile Diabetes Prevention Trial: Cross-Sectional Baseline Analysis
JMIR Diabetes 2026;11:e95295
URL: https://diabetes.jmir.org/2026/1/e95295
doi:10.2196/95295

© Nan Jiang, Jing Liu, Haili Song, Yanan Zhao, Huilin Li, Stella S Yi, Jeannette M Beasley, Lu Hu. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 3.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Emotional Burden and Behavioral Impacts of Alarms and Alerts Across the Spectrum of Diabetes Technology Use: a Cross-Sectional Study

Estelle Everett[1,2,3,4], MD, MHS; Erin Shaw[1], BS; Bryan Escobar Barrios[1], BA; Beatrice Brumley[1], MD; Kyrstin Lane[1], MD; Tannaz Moin[1,2,3,4], MD, MBA, MSHS; Lauren E Wisk[2,5], PhD

[1]Division of Endocrinology, David Geffen School of Medicine, University of California, Los Angeles, 1100 Glendon Ave, Ste 850, Los Angeles, CA, United States

[2]Division of General Internal Medicine & Health Services Research, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]VA Greater Los Angeles Healthcare System, Los Angeles, CA, United States

[4]HSR&D Center for the Study of Healthcare Innovation, Implementation, and Policy, VA Greater Los Angeles Healthcare System, Los Angeles, CA, United States

[5]Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Estelle Everett, MD, MHS
Division of Endocrinology, David Geffen School of Medicine, University of California, Los Angeles, 1100 Glendon Ave, Ste 850, Los Angeles, CA, United States

Abstract

Background: Continuous glucose monitors (CGMs), sensor-augmented pumps (SAPs), and automated insulin delivery (AID) systems have substantially improved glycemic outcomes for people with type 1 diabetes (T1D). However, these technologies also generate frequent alarms and alerts that may contribute to emotional burden, alarm fatigue, and maladaptive behavioral responses. Despite increasing recognition of alarm-related distress, little is known about how alarm burden differs across contemporary diabetes technologies.

Objective: This study evaluated differences in alarm frequency, emotional burden, and behavioral responses to alarms and alerts among adults using CGMs alone, SAP therapy, and AID systems.

Methods: We conducted a cross-sectional survey of adults (≥18 years) with T1D receiving care within a large academic health system between August 2024 and March 2025. Eligible participants completed an online survey assessing demographics, diabetes technology use, perceptions of alarm frequency and burden, and responses to alarms and alerts. Participants were categorized as CGM-only, SAP, or AID users. Differences between groups were evaluated using χ^2 tests and ANOVA. Ordinal logistic regression models adjusted for age, gender, and diabetes duration were used to examine associations between diabetes technology type and alarm-related outcomes.

Results: Among 838 respondents (mean age 46.0, SD 16.6 years; mean diabetes duration 23.0, SD 14.0 years; n=457, 55% women), AID users reported the highest alarm frequency, with 49% (n=252) experiencing alarms several times daily, compared with 35% (n=28) of SAP users and 32% (n=79) of CGM-only users (P<.001). AID users also reported greater alarm disruptiveness, annoyance, and unwanted attention than CGM-only users. Overcorrection of glucose levels in response to alerts was common across all technologies: 44% (n=371) and 50% (n=419) of participants reported sometimes overcorrecting low and high glucose levels, respectively, while 32% (n=265) and 18% (n=147) reported often or always overcorrecting low and high glucose levels, respectively. After adjustment, AID users had greater odds of frequent alarms (odds ratio [OR] 1.98, 95% CI 1.48-2.65), alarm disruptiveness (OR 1.67, 95% CI 1.24-2.25), annoyance (OR 1.65, 95% CI 1.23-2.19), unwanted attention (OR 1.85, 95% CI 1.38-2.47), ignoring alarms (OR 1.86, 95% CI 1.40-2.47), and overcorrecting low glucose levels (OR 1.61, 95% CI 1.20-2.15) compared with CGM-only users. Perceived effectiveness of alarms did not differ by technology type.

Conclusions: Although AID systems provide important clinical benefits, they are associated with the greatest alarm and alert burden. Alarm-driven behaviors, including ignoring alerts and overcorrecting glucose levels, represent previously underrecognized consequences of diabetes technology that may diminish user experience and potentially affect glycemic management. These findings identify opportunities to improve alert algorithms while highlighting the need for individualized patient education, clinician engagement in setting appropriate alert thresholds, and routine assessment of alarm burden to optimize both glycemic outcomes and patient-centered experiences with diabetes technology.

XSL•FO
RenderX

(JMIR Diabetes 2026;11:e97122) doi:10.2196/97122

KEYWORDS

insulin pump; continuous glucose monitoring; CGM; automated insulin delivery; alarm fatigue; diabetes distress; type 1 diabetes; quality of life

Introduction

Over the past 2 decades, advances in diabetes technology, particularly continuous glucose monitors (CGMs), insulin pumps, and automated insulin delivery (AID) systems, have transformed diabetes self-management. These innovations enable real-time glucose tracking, trend visualization, and algorithm-driven insulin adjustments, leading to improved glycemic outcomes and enhanced safety for many individuals with diabetes [1-5]. While diabetes technologies are clinically beneficial, they can also introduce substantial user burden [6]. One of the most pervasive sources of frustration stems from frequent device alarms and alerts. These notifications, intended to safeguard against hypo- and hyperglycemia, technical malfunctions, and connectivity issues, can be intrusive and exhausting.

Alarm and alert distress is increasingly recognized as an important contributor to the overall burden of diabetes technology use [6,7]. However, the emotional and behavioral impacts of alarms remain poorly understood. Studies assessing patient-reported outcomes have tended to use broad measures, such as the Diabetes Distress Scale, treatment satisfaction, or quality-of-life assessments, yielding inconsistent insights [6,8,9]. Few have specifically evaluated the frequency and impact of alarms and alerts, and those that have are predominantly qualitative [10-12] or limit their focus to a single technology type [13-16]. As a result, there is a significant gap in our understanding of how alarm burden differs across the full continuum of diabetes technologies, from CGM-only use to CGMs with an insulin pump in manual mode (sensor-augmented pump therapy; SAP) to the more advanced AID systems. Understanding how alarm frequency differs across diabetes technologies and how these alerts may influence emotional well-being and behavioral responses is critical for optimizing both device performance and patient experience and for facilitating shared decision-making around technology selection. The objective of this study was to compare the frequency and emotional burden of and behavioral responses to alarms and alerts among adults with type 1 diabetes (T1D) using a CGM only, SAP therapy, or an AID system. We hypothesized that increasing levels of technology automation would be associated with greater alarm burden and more frequent maladaptive behavioral responses.

Methods

Overview

We conducted a cross-sectional survey of adults (≥18 years) with T1D who used diabetes technology and received care within a large academic health system between August 2024 and March 2025. Eligible participants were identified using electronic medical record data and received invitations through their electronic patient portals. The invitation included an overview of the study and a link to an online survey administered using the Qualtrics online platform (Qualtrics LLC).

The survey included items assessing demographic (age, gender, race, ethnicity, education) and clinical characteristics (diabetes duration), current and prior use of diabetes technologies (including insulin pumps and CGMs), and perceptions of alarm and alert frequency, emotional burden, and response to alarms. Survey items were developed and refined through review of the existing literature, expert consensus, and consultation with adults living with T1D to ensure clarity and relevance. Questions were multiple choice, and responses were rated on 5-point Likert scales.

Participants were stratified by current level of diabetes technology use: CGM only, SAP, or AID systems. Participants who said they used the AID system on their pump every day or most days (5-7 days per week) were categorized as AID users; all other respondents with both a pump and a CGM were categorized as SAP users.

All analyses were conducted in SAS (version 9.4; SAS Institute Inc). All tests were 2-sided with an α criterion of .05. Demographic characteristics of the sample were compared by diabetes technology use, and appropriate bivariate tests were conducted (χ^2 and ANOVA). We used χ^2 tests to compare the prevalence of each of 9 questions about alarms and alerts by diabetes technology use. Differences in mean scores for alarm and alert questions were also compared by diabetes technology use and tested using ANOVA (Table S1). Ordinal logistic regression models were used to evaluate differences in alarm and alert scores by diabetes technology use, adjusting for age, diabetes duration, and gender (as these characteristics differed by technology use). For negatively framed outcomes, odds ratios (ORs) greater than 1.0 indicate increased odds of worse alarm and alert burden (eg, greater alert frequency, increased experience of alarm disruptiveness). For positively framed outcomes (alarm effectiveness and safety features), ORs greater than 1.0 indicate a greater endorsement of those attributes. The reporting of this cross-sectional study follows STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.

Ethical Considerations

This study was reviewed and approved by the University of California, Los Angeles Institutional Review Board (FWA00004642) on March 4, 2024 (UCLA IRB-23 - 1484). All participants provided informed consent prior to participation. A waiver of signed informed consent was granted in accordance with the Code of Federal Regulations [17]. Participants reviewed an institutional review board–approved study information statement prior to initiating the survey and provided electronic consent by selecting “yes” to participate. Participants who

XSL•FO
RenderX

selected “no” were exited from the survey and no data were collected from them.

Results

We invited 5650 adults with T1D to participate in the survey between 2024 and 2025, and a total of 1026 (18%) responded. Of these, we excluded 67 participants who did not have T1D or were younger than 18 years of age, and we further excluded 14 duplicate records. Among the 945 remaining eligible participants, we excluded those who provided incomplete answers to the CGM and pump questions (n=42), reported no diabetes technology use (n=38), or provided incomplete answers to the alarm and alert questions (n=18). Due to the small number of insulin pump–only users (n=9), these participants were also excluded from the analysis. The final analytic sample comprised 838 eligible participants (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of participants included in the final analytic sample. CGM: continuous glucose monitor.

XSL•FO
RenderX

The characteristics of the participants are reported in Table 1. Overall, participants' mean age was 45.76 (SD 16.91) years, and mean diabetes duration was 22.84 (SD 13.96) years. Mean age differed significantly across groups; AID users were younger (mean age 43.86, SD 16.03 years) than SAP users (mean age 50.13, SD 18.95 years) and CGM-only users (mean age 48.36, SD 17.47 years; $P<.001$). CGM-only and AID users had shorter mean diabetes durations of 20.11 (SD 15.01) years and 23.59 (SD 13.20) years, respectively, compared with SAP users (mean 26.39, SD 14.20 years; $P<.001$). Men were more likely to be CGM-only users (n=125, 51%; $P<.001$). There were no significant differences in type of technology used by race and ethnicity or educational attainment. Mean duration of CGM use was 5.10 (SD 4.20) years among CGM-only users and longer among SAP (mean duration 8.14, SD 5.58 years) and AID users (mean duration 7.99, SD 4.78 years; $P<.001$). Mean duration of insulin pump use did not differ significantly between SAP and AID users (12.42, SD 8.71 vs 12.52, SD 8.46 years; $P=.92$).

XSL•FO
RenderX

Table . Characteristics of participants.

Variables	All (n=838)	CGM[a] only (n=244)	CGM and pump—manual mode (SAP[b]; n=79)	CGM and pump—automated mode (AID[c]; n=515)	*P* value
Age (years), mean (SD)	45.76 (16.91)	48.36 (17.47)	50.13 (18.95)	43.86 (16.03)	<.001
Age by subgroup (years), n (%)					<.001
18 - 29	164 (20)	46 (19)	16 (20)	102 (20)	
30 - 39	206 (25)	39 (16)	15 (19)	152 (29)	
40 - 49	135 (16)	44 (18)	10 (13)	81 (16)	
50 - 59	120 (14)	37 (15)	10 (13)	73 (14)	
60 - 69	126 (15)	45 (18)	13 (16)	68 (13)	
≥70	87 (10)	33 (14)	15 (19)	39 (8)	
Diabetes duration (years), mean (SD)	22.84 (13.96)	20.11 (15.01)	26.39 (14.20)	23.59 (13.20)	<.001
Diabetes duration by subgroup (years), n (%)					<.001
0 - 5	89 (11)	47 (19)	4 (5)	38 (7)	
6 - 10	98 (12)	41 (17)	6 (8)	51 (10)	
11 - 15	110 (13)	22 (9)	10 (13)	78 (15)	
16 - 20	117 (14)	33 (14)	14 (18)	70 (14)	
21 - 25	108 (13)	25 (10)	8 (10)	75 (15)	
26 - 30	78 (9)	18 (7)	11 (13)	49 (9)	
≥31	238 (28)	58 (24)	26 (33)	154 (30)	
Gender, n (%)					<.001
Women	457 (55)	108 (44)	42 (53)	307 (60)	
Men	339 (40)	125 (51)	32 (41)	182 (35)	
Other/missing	42 (5)	11 (5)	5 (6)	26 (5)	
Race and ethnicity, n (%)					.61
Hispanic	108 (13)	32 (13)	5 (6)	71 (14)	
Non-White, non-Hispanic	99 (12)	31 (13)	12 (15)	56 (11)	
White, non-Hispanic	586 (70)	168 (69)	57 (73)	361 (70)	
Missing	45 (5)	13 (5)	5 (6)	27 (5)	
Education, n (%)					
High school diploma or GED[d] or less	27 (3)	13 (5)	2 (3)	12 (2)	.35
Some college (no degree)	120 (14)	39 (16)	11 (14)	70 (14)	
Associate's degree	55 (7)	19 (8)	5 (6)	31 (6)	
Bachelor's degree	340 (41)	104 (42)	28 (35)	208 (40)	
Master's degree	183 (22)	43 (18)	22 (28)	118 (23)	
PhD, MD, or other advanced degree	79 (9)	17 (7)	7 (9)	55 (11)	
Prefer not to say/missing	34 (4)	9 (4)	4 (5)	21 (4)	

XSL•FO
RenderX

Variables	All (n=838)	CGM[a] only (n=244)	CGM and pump—manual mode (SAP[b]; n=79)	CGM and pump—automated mode (AID[c]; n=515)	*P* value
Duration of CGM use (years), mean (SD)	7.16 (4.88)	5.10 (4.20)	8.14 (5.58)	7.99 (4.78)	<.001
Duration of CGM use by subgroup (years), n (%)					<.001
0 - 5	367 (44)	158 (65)	29 (37)	180 (35)	
6 - 10	335 (40)	68 (28)	35 (44)	232 (45)	
≥11	136 (16)	18 (7)	15 (19)	103 (20)	
Duration of pump use[e] (years), mean (SD)	12.51(8.49)	—[f]	12.42 (8.71)	12.52 (8.46)	.92
Duration of pump use by subgroup (years), n (%)					.92
0 - 5	167 (28)	—	23 (29)	144 (28)	
6 - 10	114 (19)	—	18 (23)	96 (19)	
11 - 15	108 (18)	—	9 (11)	99 (19)	
16 - 20	90 (15)	—	15 (19)	75 (14)	
≥21	115 (20)	—	14 (18)	101 (20)	

[a]CGM: continuous glucose monitor.

[b]SAP: sensor-augmented pump.

[c]AID: automated insulin delivery.

[d]GED: General Education Development.

[e]Participants who use any type of pump, n=594.

[f]Not applicable.

The frequency of alarms and alerts differed significantly across technology groups (*P*<.001; Table 2). AID users reported the highest frequency, with 49% (n=252) experiencing alarms or alerts several times per day, compared with 35% (n=28) of SAP users and 32% (n=79) of CGM-only users. The psychological impact of these alarms was also greatest among insulin pump users: 32% (n=165) of AID and 30% (n=24) of SAP users reported that alarms and alerts were often or always disruptive compared with 21% (n=50) of CGM-only users (*P*=.009). Similarly, 41% (n=207) of AID users and 37% (n=29) of SAP users found alarms and alerts often or always annoying versus 28% (n=70) of CGM-only users. Alarms and alerts were also more likely to draw unwanted attention among pump users, with 21% (n=111) of AID and 24% (n=19) of SAP users reporting this often or always compared with 14% (n=34) of CGM-only users.

XSL•FO
RenderX

Table . Differences in alarm and alert (AA) frequency, burden, and responses by diabetes technology.

Variables	All (n=838), n (%)	CGM[a] only (n=244), n (%)	CGM and pump—manual mode (SAP[b]; n=79), n (%)	CGM and pump—automated mode (AID[c]; n=515), n (%)	*P* value
AA frequency					<.001
Several times a day	359 (43)	79 (32)	28 (35)	252 (49)	
About once a day	246 (29)	67 (28)	21 (27)	158 (31)	
Every few days	156 (19)	65 (27)	18 (23)	73 (14)	
About once a week	49 (6)	18 (7)	3 (4)	28 (5)	
Never	28 (3)	15 (6)	9 (11)	4 (1)	
AA disruptiveness					.009
Never	36 (4)	16 (7)	6 (8)	14 (3)	
Rarely	163 (19)	56 (23)	18 (23)	89 (17)	
Sometimes	400 (48)	122 (49)	31 (39)	247 (48)	
Often	208 (25)	44 (18)	21 (26)	143 (28)	
Always	31 (4)	6 (3)	3 (4)	22 (4)	
AA unwanted attention					<.001
Never	97 (12)	47 (19)	10 (13)	40 (8)	
Rarely	247 (29)	78 (32)	26 (33)	143 (28)	
Sometimes	330 (39)	85 (35)	24 (30)	221 (43)	
Often	134 (16)	27 (11)	17 (21)	90 (17)	
Always	30 (4)	7 (3)	2 (3)	21 (4)	
AA annoyance					<.001
Never	60 (7)	28 (12)	10 (13)	22 (4)	
Rarely	152 (18)	52 (21)	12 (15)	88 (17)	
Sometimes	320 (38)	94 (39)	28 (35)	198 (38)	
Often	219 (26)	57 (23)	21 (27)	141 (28)	
Always	87 (11)	13 (5)	8 (10)	66 (13)	
AA effectiveness					.27
Never	14 (2)	7 (3)	1 (1)	6 (1)	
Rarely	79 (9)	22 (9)	7 (9)	50 (10)	
Sometimes	257 (31)	84 (34)	25 (32)	148 (29)	
Often	316 (38)	77 (32)	34 (43)	205 (40)	
Always	172 (20)	54 (22)	12 (15)	106 (20)	
AA safety feature					.03
Never	11 (1)	5 (2)	1 (1)	5 (1)	
Rarely	57 (7)	17 (7)	10 (13)	30 (6)	
Sometimes	176 (21)	36 (15)	13 (16)	127 (25)	
Often	255 (30)	76 (31)	22 (28)	157 (30)	
Always	339 (41)	110 (45)	33 (42)	196 (38)	
Overcorrecting highs					.21
Never	38 (4)	14 (6)	7 (9)	17 (3)	
Rarely	234 (28)	58 (24)	21 (26)	155 (30)	
Sometimes	419 (50)	123 (50)	41 (52)	255 (50)	

XSL•FO
RenderX

Variables	All (n=838), n (%)	CGM[a] only (n=244), n (%)	CGM and pump—manual mode (SAP[b]; n=79), n (%)	CGM and pump—automated mode (AID[c]; n=515), n (%)	*P* value
Often	140 (17)	47 (19)	10 (13)	83 (16)	
Always	7 (1)	2 (1)	0 (0)	5 (1)	
Overcorrecting lows					.03
Never	34 (4)	12 (5)	7 (9)	15 (3)	
Rarely	168 (20)	62 (26)	16 (20)	90 (17)	
Sometimes	371 (44)	106 (43)	30 (38)	235 (46)	
Often	216 (26)	54 (22)	20 (25)	142 (28)	
Always	49 (6)	10 (4)	6 (8)	33 (6)	
Ignoring AA					<.001
Never	111 (13)	40 (17)	18 (23)	53 (10)	
Rarely	229 (27)	86 (35)	14 (18)	129 (25)	
Sometimes	240 (29)	69 (28)	25 (32)	146 (28)	
Often	218 (26)	42 (17)	17 (21)	159 (31)	
Always	40 (5)	7 (3)	5 (6)	28 (6)	

[a]CGM: continuous glucose monitor.

[b]SAP: sensor-augmented pump.

[c]AID: automated insulin delivery.

Responses to alarms and alerts differed across technology groups. Over one-third of AID users (n=258, 37%) reported often or always ignoring alarms compared with 27% (n=22) of SAP users and 20% (n=49) of CGM-only users (*P*<.001). Approximately one-third of both AID (n=175, 34%) and SAP (n=26, 33%) users reported often or always overcorrecting low glucose levels by dosing too little insulin or treating low blood glucose with food instead of adjusting insulin in response to alerts compared with 26% (n=64) of CGM-only users (*P*=.03). In contrast, overcorrection of high glucose levels often or always by dosing too much insulin in response to alerts was reported more frequently among CGM-only users (n=49, 20%) than among AID (n=88, 17%) and SAP users (n=10, 13%).

Despite these differences in behavioral responses to alarms, most participants found alarms and alerts to be effective, with 60% (n=311) of AID, 58% (n=46) of SAP, and 54% (n=131) of CGM-only users reporting that alarms and alerts were often or always helpful (*P*=.27). However, CGM-only users were more likely to view alarms and alerts as an important safety feature (76%) compared with SAP (70%) and AID users (68%; *P*=.03).

In adjusted models (Table 3), AID users reported a higher frequency of alarms and alerts compared with CGM-only users (OR 1.98, 95% CI 1.48 - 2.65), as well as greater perceived disruptiveness (OR 1.67, 95% CI 1.24 - 2.25), annoyance (OR 1.65, 95% CI 1.23 - 2.19), and unwanted attention (OR 1.85, 95% CI 1.38 - 2.47). There were no significant differences in perceived effectiveness (OR 0.80, 95% CI 0.60 - 1.07) or in viewing alarms and alerts as an important safety feature (OR 1.31, 95% CI 0.98 - 1.76). AID users were also more likely to ignore alarms (OR 1.86, 95% CI 1.40 - 2.47) and overcorrect low glucose levels (OR 1.61, 95% CI 1.20 - 2.15). No differences were observed in overcorrection of high glucose levels (OR 0.85, 95% CI 0.63 - 1.15).

XSL•FO
RenderX

Table . Adjusted associations between diabetes technology type and alarm and alert (AA) frequency, burden, and responses.

Variables	AA frequency, OR[a] (95% CI)	AA disruptiveness, OR (95% CI)	AA unwanted attention, OR (95% CI)	AA annoyance, OR (95% CI)	AA effectiveness, OR (95% CI)	AA safety feature, OR (95% CI)	Overcorrecting highs, OR (95% CI)	Overcorrecting lows, OR (95% CI)	Ignoring AAs, OR (95% CI)
Device/tech use									
SAP[b] vs CGM[c]	1.03 (0.65 - 1.64)	1.27 (0.79 - 2.05)	1.65 (1.03 - 2.64)	1.35 (0.85 - 2.15)	0.97 (0.61-1.55)	1.37 (0.86-2.20)	0.69 (0.42-1.11)	1.24 (0.78-1.99)	1.45 (0.91-2.29)
AID[d] vs CGM	1.98 (1.48 - 2.65)	1.67 (1.24 - 2.25)	1.85 (1.38 - 2.47)	1.65 (1.23-2.19)	0.80 (0.60-1.07)	1.31 (0.98-1.76)	0.85 (0.63-1.15)	1.61 (1.20-2.15)	1.86 (1.40-2.47)
Age	0.98 (0.97 - 0.99)	0.99 (0.99 - 1.00)	0.99 (0.98 - 1.00)	0.98 (0.97-0.99)	0.99 (0.98-0.99)	0.98 (0.98-0.99)	0.99 (0.98-1.00)	1.00 (0.99-1.01)	0.97 (0.96-0.98)
Diabetes duration	1.01 (1.00 - 1.02)	0.99 (0.99 - 1.00)	0.99 (0.98 - 1.00)	1.00 (0.99-1.01)	1.02 (1.01-1.03)	1.00 (0.99-1.01)	1.01 (1.00-1.02)	0.99 (0.98-1.00)	1.00 (0.99-1.01)
Gender									
Men vs women	0.92 (0.71 - 1.20)	0.69 (0.53 - 0.90)	0.69 (0.53 - 0.89)	0.67 (0.52-0.87)	1.13 (0.87-1.46)	1.50 (1.16-1.96)	1.35 (1.03-1.77)	1.06 (0.81-1.38)	1.08 (0.84-1.40)
Other vs women	1.08 (0.60 - 1.95)	1.17 (0.65 - 2.10)	0.74 (0.42 - 1.33)	1.44 (0.81-2.56)	0.66 (0.37-1.18)	0.88 (0.49-1.59)	1.29 (0.71-2.34)	1.34 (0.75-2.39)	1.29 (0.73-2.29)

[a]OR: odds ratio.

[b]SAP: sensor-augmented pump.

[c]CGM: continuous glucose monitor.

[d]AID: automated insulin delivery.

In adjusted models comparing SAP to CGM-only users, SAP users reported greater feelings of unwanted attention (OR 1.65, 95% CI 1.03 - 2.64) but did not differ significantly in alarm frequency, disruptiveness, annoyance, perceived effectiveness, safety perception, or responses to alarms (ignoring or overcorrecting).

Alarm responses varied by demographic characteristics. Although there were no gender differences in reported alarm frequency (OR 0.92, 95% CI 0.71 - 1.20), men were less likely than women to perceive alarms as disruptive (OR 0.69, 95% CI 0.53 - 0.90) or to report that alarms drew unwanted attention (OR 0.69, 95% CI 0.53 - 0.89). Men were more likely to view alarms as an important safety feature (OR 1.50, 95% CI 1.16 - 1.96) and were more likely to overcorrect high glucose levels in response to alerts (OR 1.35, 95% CI 1.03 - 1.77).

Discussion

In this cross-sectional study of adults with T1D using the spectrum of contemporary diabetes technologies, we compared alarm frequency, emotional burden, and behavioral responses across CGM-only, SAP, and AID system users. We found that AID users experienced the greatest alarm burden, reporting more frequent and disruptive alarms and greater maladaptive behavioral responses to alarms than users of less automated technologies. These findings highlight an important trade-off between the clinical benefits of automation and the psychosocial burden associated with increasing device complexity.

Several factors likely contribute to the greater alarm burden observed among AID users. The combination of insulin pump and CGM use may increase alarm burden because each device can generate alerts through separate safety and monitoring mechanisms. Common causes of insulin pump alerts include obstruction of insulin flow, infusion set or site problems, and device or reservoir issues [11,18]. Additionally, a CGM may issue alerts due to true glycemic excursions, but such alerts can also result from sensor inaccuracies, compression lows (eg, due to sleep position), loss of connectivity, or device malfunction [19,20].

The heightened alarm burden among AID users likely reflects both the integrated nature of these systems and their continuous feedback mechanisms. AID systems include several unique alarms and alerts tied to their control algorithms and safety features, including alerts when the algorithm cannot safely determine insulin delivery, mode transition alarms when switching between automated and manual operation, and notifications for automated correction boluses. AID systems also issue predictive safety alarms for impending hypo- or hyperglycemia and alerts for potential system failures or overrides. In addition, many contemporary insulin pumps include integrated speakers and can generate audible glucose-level alerts based on connected CGM data, while the same glucose event may simultaneously trigger alerts on a smartphone, dedicated receiver, or smartwatch. Consequently, users may receive multiple alerts for a single glucose excursion, each requiring individual acknowledgment or dismissal, further amplifying alarm burden. While these alerts are critical for safety and algorithm performance, the sheer frequency may contribute to alarm fatigue and reduced responsiveness over time. Our finding that AID users were more likely to ignore

XSL•FO
RenderX

alarms underscores this concern and parallels alarm fatigue phenomena reported in other health care settings, such as bedside monitoring [21-23].

Our findings also suggest that frequent alarms and alerts may lead to maladaptive behavioral responses, including overcorrection in response to both low and high glucose–level alerts. Across our cohort of 838 respondents, 44% (n=371) and 50% (n=419) of participants reported sometimes overcorrecting for low and high glucose levels, respectively, and 32% (n=265) and 18% (n=147) reported often or always overcorrecting for low and high glucose levels, respectively, in response to alerts. This represents a novel phenomenon that has not been well characterized in prior diabetes technology research. These behaviors may reflect heightened vigilance or anxiety triggered by frequent alerts, prompting users to respond preemptively or excessively to perceived glucose excursions. While glycemic alarms are intended to promote safety, repeated overcorrection may paradoxically increase glucose variability and contribute to a “roller-coaster” pattern of glycemia, with rapid swings from high to low or vice versa. Further investigation is warranted to quantify how often overcorrection precipitates such fluctuations and to understand its downstream impact on key glycemic outcomes, including hemoglobin A_{1c} level, time in range, and hypoglycemia risk, particularly among users of AID systems.

Psychological and behavioral responses to alarms varied by gender. Women reported higher levels of disruptiveness and unwanted attention, consistent with prior research showing that female [10,24-27] users may experience greater diabetes-related burden and social visibility concerns. Men were more likely to view alarms as a safety feature and more likely to overcorrect highs, which may reflect differences in risk perception and response behaviors by gender [28,29]. These patterns highlight the need for tailored education and support that address demographic and psychosocial differences in alarm experiences.

Our study adds to the growing literature emphasizing the importance of user-centered outcomes in diabetes technology evaluation. While glycemic metrics remain the gold standard for assessing device efficacy, psychosocial factors, including alarm fatigue, distress, and technology satisfaction, strongly influence sustained device use and quality of life. Our findings suggest that alarm burden should be considered alongside glycemic outcomes when selecting, initiating, and optimizing diabetes technologies. Although AID and SAP users experienced greater alarm-related distress, perceived effectiveness and safety of alarms were similar across device types, suggesting that users recognize the clinical value of alerts despite their inconvenience. This coexistence of perceived utility and burden highlights a central design challenge: reducing unnecessary alerts while preserving safety-critical notifications. Clinicians should routinely assess alarm-related distress during diabetes visits and work with patients to personalize alarm settings, optimize notification thresholds, and address maladaptive responses such as alarm avoidance and overcorrection. These findings also underscore the importance of ongoing patient education to help individuals understand the purpose of alerts, develop appropriate responses, and safely manage frequent notifications without disengaging from their devices. Patients experiencing frequent alarms may benefit from periodic reassessment of alert settings and additional clinical support to ensure that alarms remain both meaningful and actionable. Likewise, manufacturers should continue to refine algorithms, personalize notification thresholds, and apply user-centered design principles to develop alert systems that are responsive, flexible, and minimally intrusive.

The limitations of this study include its cross-sectional design, which precludes causal inference. Participants were drawn from a single academic health system, which may limit generalizability. Self-reported measures may also be subject to recall or response bias. Additionally, we used a study-specific survey rather than a standardized, validated patient-reported outcome measure. While we assessed perceived burden and behavioral responses, we did not directly measure alarm frequency or glycemic data using device downloads. We also did not evaluate differences in alarm burden across individual CGM or insulin pump systems; however, device-specific features, including differences in alert modalities, may influence users’ experiences and should be examined in future studies.

This study also has several strengths, including its large sample, literature-informed survey instrument, and inclusion of users across the full spectrum of contemporary diabetes technologies. To our knowledge, this is the first study to quantitatively compare alarm frequency, burden, and behavioral responses across these technologies. Future studies incorporating objective device data and longitudinal follow-up will be important to determine how alarm burden influences glycemic outcomes and long-term technology engagement.

In conclusion, as diabetes technologies become increasingly automated, optimizing the user experience must become a priority alongside improving glycemic outcomes. Addressing alarm-related distress and its psychological and behavioral consequences is essential to ensure that advances in automation translate into meaningful improvements in quality of life, sustained technology engagement, and clinical outcomes. A user-centered approach to device design, individualized patient education, and proactive clinical management, including routine review of alarm settings and patient responses to alerts, will be critical to realizing the full potential of diabetes technologies.

Acknowledgments

While preparing this work, the authors used ChatGPT for the purpose of proofreading, refining written text to improve clarity, and ensuring concise language. Following the use of this tool, the authors formally reviewed the content for its accuracy and edited it as necessary. The authors take full responsibility for all of the content of this publication.

XSL•FO
RenderX

Funding

EE received support from the National Institutes of Health (NIH)/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (K23DK132482 and L40DK129996) and the Harold Amos Medical Faculty Development Program. TM also received support from the NIH/NIDDK (R01DK124503 and R01DK127733), Centers for Disease Control and Prevention (U18 DP006708), University of California, Los Angeles, University of California Office of the President, and Department of Veterans Affairs (CSP#2002 and CSP NODES0017). LEW received support from the American Diabetes Association (7-23-ICTST2DY-11), National Institute of Mental Health (R01MH131597), and NIDDK (R01DK144220). This research was supported by the David Geffen School of Medicine's Justice, Equity, Diversity and Inclusion Junior Faculty Academic Mentoring Council Program Seed Funding.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional policies and the inclusion of potentially identifiable participant information but are available from the corresponding author on reasonable request.

Authors' Contributions

LEW and EE contributed to study conceptualization, methodology, and funding acquisition. LEW performed the formal analysis and validation. EE drafted the original manuscript. BEB managed project administration. BB contributed to methodology and project administration. ES contributed to project administration, visualization, and manuscript review and editing. TM provided supervision and contributed to manuscript review and editing. KL contributed to methodology and manuscript review and editing. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Teo E, Hassan N, Tam W, Koh S. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. Diabetologia 2022 Apr;65(4):604-619. [doi: 10.1007/s00125-021-05648-4] [Medline: 35141761]
2. Yeh HC, Brown TT, Maruthur N, et al. Comparative effectiveness and safety of methods of insulin delivery and glucose monitoring for diabetes mellitus: a systematic review and meta-analysis. Ann Intern Med 2012 Sep 4;157(5):336-347. [doi: 10.7326/0003-4819-157-5-201209040-00508] [Medline: 22777524]
3. Di Molfetta S, Di Gioia L, Caruso I, et al. Efficacy and safety of different hybrid closed loop systems for automated insulin delivery in people with type 1 diabetes: a systematic review and network meta-analysis. Diabetes Metab Res Rev 2024 Sep;40(6):e3842. [doi: 10.1002/dmrr.3842] [Medline: 39298688]
4. Yang Q, Zeng B, Hao J, Yang Q, Sun F. Real-world glycaemic outcomes of automated insulin delivery in type 1 diabetes: a meta-analysis. Diabetes Obes Metab 2024 Sep;26(9):3753-3763. [doi: 10.1111/dom.15718] [Medline: 38888056]
5. Pöhlmann J, Smith-Palmer J, Serné EH, et al. A systematic literature review and meta-analysis of real-world evidence on commercially available automated insulin delivery systems in people with type 1 diabetes. Diabetes Obes Metab 2025 Oct 8;28 Suppl 1(Suppl 1):13-34. [doi: 10.1111/dom.70161] [Medline: 41058520]
6. Barnard-Kelly K, Gonder-Frederick L, Weissberg-Benchell J, Wisk LE. Psychosocial aspects of diabetes technologies: commentary on the current status of the evidence and suggestions for future directions. J Diabetes Sci Technol 2025 Jan;19(1):27-33. [doi: 10.1177/19322968241276550] [Medline: 39431295]
7. Polonsky WH. Psychosocial aspects of diabetes technology: adult perspective. Endocrinol Metab Clin North Am 2020 Mar;49(1):143-155. [doi: 10.1016/j.ecl.2019.10.003] [Medline: 31980114]
8. Roos T, Hermanns N, Groß C, Kulzer B, Haak T, Ehrmann D. Effect of automated insulin delivery systems on person-reported outcomes in people with diabetes: a systematic review and meta-analysis. EClinicalMedicine 2024 Oct;76:102852. [doi: 10.1016/j.eclinm.2024.102852] [Medline: 39364272]
9. Gao M, Saravanan S, Munyombwe T, et al. Effects of modern glucose monitoring and insulin delivery technologies on patient-reported outcomes and experiences in individuals with type 1 diabetes: a systematic review and meta-analysis. Diabetes Care 2025 Dec 1;48(12):2160-2180. [doi: 10.2337/dc25-1178] [Medline: 41166569]
10. McKechnie V, Avari P, Eng PC, et al. The experiences of high-risk young adults with type 1 diabetes transitioning to real-time continuous glucose monitoring—a thematic analysis. PLoS ONE 2025;20(5):e0320702. [doi: 10.1371/journal.pone.0320702] [Medline: 40315203]
11. Estock JL, Codario RA, Zupa MF, Keddem S, Rodriguez KL. Insulin pump alarms during adverse events: a qualitative descriptive study. J Diabetes Sci Technol 2025 May;19(3):739-748. [doi: 10.1177/19322968231209999] [Medline: 37905930]
12. Miller E, Midyett LK. Just because you can, doesn't mean you should...now. A practical approach to counseling persons with diabetes on use of optional CGM alarms. Diabetes Technol Ther 2021 Sep;23(S3):S66-S71. [doi: 10.1089/dia.2021.0192] [Medline: 34546081]

XSL•FO
RenderX

13. Grando MA, Bayuk M, Karway G, et al. Patient perception and satisfaction with insulin pump system: pilot user experience survey. J Diabetes Sci Technol 2019 Nov;13(6):1142-1148. [doi: 10.1177/1932296819843146] [Medline: 31055947]
14. Wu Z, Alexandre-Heymann L, Lebbar M, et al. Despite lower haemoglobin A1c with second-generation automated insulin delivery systems, mental burden remains high for all adults with type 1 diabetes: A BETTER registry analysis. Diabetes Obes Metab 2026 Feb;28(2):1150-1159. [doi: 10.1111/dom.70299] [Medline: 41250885]
15. González-Vidal T, Rivas-Otero D, Agüeria-Cabal P, Ramos-Ruiz G, Delgado E, Menéndez-Torre E. Continuous glucose monitoring alarms in adults with type 1 diabetes: user characteristics and the impact of hypoglycemia and hyperglycemia alarm thresholds on glycemic control. Diabetes Technol Ther 2024 May;26(5):313-323. [doi: 10.1089/dia.2023.0460] [Medline: 38156962]
16. Cobry EC, Vigers T, Berget C, et al. Frequency and causes of nocturnal alarms in youth and young adults with type 1 diabetes using a first-generation hybrid closed-loop system. Diabetes Spectr 2024;37(2):118-123. [doi: 10.2337/ds23-0003] [Medline: 38756430]
17. Documentation of informed consent. 42 CFR §447.502. eCFR. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.117 [accessed 2026-07-16]
18. Estock JL, Codario RA, Keddem S, Zupa MF, Rodriguez KL, DiNardo MM. Insulin pump-associated adverse events: a qualitative descriptive study of clinical consequences and potential root causes. Diabetes Technol Ther 2023 May;25(5):343-355. [doi: 10.1089/dia.2022.0498] [Medline: 36724310]
19. Sherr JL, Heinemann L, Fleming GA, et al. Automated insulin delivery: benefits, challenges, and recommendations. A consensus report of the joint diabetes technology working group of the European Association for the Study of Diabetes and the American Diabetes Association. Diabetes Care 2022 Dec 1;45(12):3058-3074. [doi: 10.2337/dci22-0018] [Medline: 36202061]
20. Pleus S, Eichenlaub M, Waldenmaier D, Freckmann G. A critical discussion of alert evaluations in the context of continuous glucose monitoring system performance. J Diabetes Sci Technol 2024 Jul;18(4):847-856. [doi: 10.1177/19322968241236504] [Medline: 38477308]
21. Lewandowska K, Weisbrot M, Cieloszyk A, Mędrzycka-Dąbrowska W, Krupa S, Ozga D. Impact of alarm fatigue on the work of nurses in an intensive care environment—a systematic review. Int J Environ Res Public Health 2020 Nov 13;17(22):1-14. [doi: 10.3390/ijerph17228409] [Medline: 33202907]
22. Hravnak M, Pellathy T, Chen L, et al. A call to alarms: current state and future directions in the battle against alarm fatigue. J Electrocardiol 2018;51(6S):S44-S48. [doi: 10.1016/j.jelectrocard.2018.07.024] [Medline: 30077422]
23. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. Curr Opin Anaesthesiol 2015 Dec;28(6):685-690. [doi: 10.1097/ACO.0000000000000260] [Medline: 26539788]
24. Tanenbaum ML, Hanes SJ, Miller KM, Naranjo D, Bensen R, Hood KK. Diabetes device use in adults with type 1 diabetes: barriers to uptake and potential intervention targets. Diabetes Care 2017 Feb;40(2):181-187. [doi: 10.2337/dc16-1536] [Medline: 27899489]
25. Messer LH, Tanenbaum ML, Cook PF, et al. Cost, hassle, and on-body experience: barriers to diabetes device use in adolescents and potential intervention targets. Diabetes Technol Ther 2020 Oct;22(10):760-767. [doi: 10.1089/dia.2019.0509] [Medline: 32163719]
26. Lorenzen JT, Madsen KP, Cleal B, et al. Associations between use of diabetes technology and diabetes distress: a Danish cross-sectional survey of adults with type 1 diabetes. BMJ Open 2024 Mar 25;14(3):e080053. [doi: 10.1136/bmjopen-2023-080053] [Medline: 38531585]
27. Smith MB, Albanese-O'Neill A, Macieira TGR, et al. Human factors associated with continuous glucose monitor use in patients with diabetes: a systematic review. Diabetes Technol Ther 2019 Oct;21(10):589-601. [doi: 10.1089/dia.2019.0136] [Medline: 31335196]
28. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, et al. Sex and gender: modifiers of health, disease, and medicine. Lancet 2020 Aug;396(10250):565-582. [doi: 10.1016/S0140-6736(20)31561-0] [Medline: 32828189]
29. Gustafson PE. Gender differences in risk perception: theoretical and methodological perspectives. Risk Anal 1998 Dec;18(6):805-811. [doi: 10.1023/b:rian.0000005926.03250.c0] [Medline: 9972583]

Abbreviations

AID: automated insulin delivery
CGM: continuous glucose monitor
OR: odds ratio
SAP: sensor-augmented pump
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
T1D: type 1 diabetes

XSL•FO
RenderX

Edited by I Steenstra; submitted 03.Apr.2026; peer-reviewed by A Welch, J Layne; revised version received 02.Jul.2026; accepted 03.Jul.2026; published 13.Aug.2026.

Please cite as:
Everett E, Shaw E, Escobar Barrios B, Brumley B, Lane K, Moin T, Wisk LE
The Emotional Burden and Behavioral Impacts of Alarms and Alerts Across the Spectrum of Diabetes Technology Use: a Cross-Sectional Study
JMIR Diabetes 2026;11:e97122
URL: https://diabetes.jmir.org/2026/1/e97122
doi:10.2196/97122

© Estelle Everett, Erin Shaw, Bryan Escobar Barrios, Beatrice Brumley, Kyrstin Lane, Tannaz Moin, Lauren E Wisk. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 13.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the REACHOUT Mental Health Support App for Type 1 Diabetes From the Perspectives of Recipients and Providers of Peer Support: Qualitative Study

Debbie Lam[1], MSc; Poonamdeep Jhajj[2], MD; Diana Sherifali[3,4,5], RN, CDE, PhD; Frances S Chen[6], PhD; Tricia S Tang[7], PhD, RPsych

[1]Experimental Medicine Program, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]Research Institute, BC Children's Hospital, Vancouver, BC, Canada
[3]School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[4]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
[5]Heather M. Arthur Population Health Research Institute/Hamilton Health Sciences Chair in Inter-Professional Health Research, Hamilton, ON, Canada
[6]Department of Psychology, Faculty of Arts, University of British Columbia, Vancouver, BC, Canada
[7]Division of Endocrinology, Faculty of Medicine, University of British Columbia, 2775 Laurel Street, Vancouver, BC, Canada

Corresponding Author:
Tricia S Tang, PhD, RPsych
Division of Endocrinology, Faculty of Medicine, University of British Columbia, 2775 Laurel Street, Vancouver, BC, Canada

Abstract

Background: Existing qualitative research in peer support interventions has largely focused on the recipients of support rather than those delivering support. Exploring the perspectives of both roles may provide a holistic understanding of the peer support experience.

Objective: This study elicits the experiences of recipients and providers of support who participated in REACHOUT, a 6-month peer-led mental health support intervention delivered via mobile app for adults with type 1 diabetes. REACHOUT offered multiple support delivery modalities (one-on-one, group-based texting, and virtual face-to-face small group sessions) that could be customized by recipients.

Methods: A total of 32 study participants (recipients and peer supporters) attended focus group discussions following the completion of REACHOUT. Thematic analysis was performed in an inductive approach.

Results: Four major themes were identified by thematic analysis: (1) need for a sense of community and belonging, (2) factors to enhance the recipient-peer supporter experience, (3) key aspects of the peer supporter experience, and (4) importance of personalizing the user experience while using the REACHOUT mobile app. REACHOUT successfully fostered connectedness by bringing together adults with type 1 diabetes who previously felt isolated. Recipients felt greater agency when given the opportunity to self-select a peer supporter. The main factors considered during the matching process included insulin delivery and glucose monitoring systems, duration of diabetes, shared hobbies, life stage, and age. While support was designed to be unidirectional from peer supporter to recipient, the former also derived benefits. Peer supporters expressed the need for greater guidance around navigating boundaries and responding to emotionally charged conversations. Finally, the REACHOUT app was able to accommodate a heterogeneity of support needs by offering one-on-one and group support across multiple communication platforms including text, audio, and video.

Conclusions: The success of peer-led mental health support interventions such as REACHOUT is likely associated with the recipient-peer supporter dynamic. By offering a range of support delivery and communication modalities, participants can better personalize solutions to meet their unique support needs. Understanding the perspectives of both recipients and peer supporters is essential to refining interventions and optimizing digitally delivered mental health support models.

(JMIR Diabetes 2026;11:e72779) doi:10.2196/72779

KEYWORDS

diabetes; mental health; mHealth; mobile app; mobile health; peer support; qualitative; thematic analysis; type 1; type 1 diabetes

XSL•FO
RenderX

Introduction

Peer support is a promising self-management strategy to improve emotional health in chronic illness care [1-4]. In the context of diabetes, several systematic reviews of adults with diabetes (both type 1 and type 2 diabetes) have found peer support interventions to be associated with improved clinical, behavioral, and psychosocial (quality of life, perceived social support) outcomes [5-8]. However, to better understand the processes underlying these positive changes, it is important to explore the qualitative experience of giving and receiving peer support.

While qualitative research on peer support interventions has focused largely on the experiences of those who receive support [9-11], there has been a notable increase in studies focused on the individuals who deliver support [12-19]. However, the optimal model for understanding the peer support experience is to explore the perspectives of both parties involved. To date, there have been 4 qualitative studies that have investigated the experiences of both recipients of support and peer supporters in the context of diabetes [20-23]. Of these studies, only 1 recruited adults with type 1 diabetes (T1D) as part of a larger sample [21], while the other investigations targeted adults with type 2 diabetes [20,22,23].

In the era of digital health, peer support models in diabetes have been made more accessible through the shift to virtual platforms such as mobile apps. Such digital peer support programs are especially valuable in rural and remote areas, where access to traditional peer networks and diabetes programs can be limited [24-26]. A systematic review of in-person and technology-mediated peer support for adults with diabetes found that peer support was beneficial in reducing isolation and increasing social support for recipients [27]. However, none of these studies were specific to T1D only. Interestingly, in a review of technology for peer support intervention for adolescents with chronic illness, rather than adults, T1D was the most represented condition [28]. Generally, adolescents with T1D experienced benefits in emotional support and diabetes management [29]. Of the few studies utilizing mobile or web apps for T1D adults, peer support was a secondary feature to self-management behavior education or one of multiple intervention components rather than the main focus [30-33]. As T1D is a lifelong condition, it is important to offer ongoing mental health support to adults living with T1D, especially those facing geographical or resource barriers.

Methods

Study Aim

This study aimed to explore the experiences of and perspectives from recipients and providers of support on REACHOUT, a peer-led mental health support intervention for adults with T1D living in rural and remote regions of British Columbia, Canada.

Study Design

Following the completion of the pilot trial titled REACHOUT, which investigated the feasibility and acceptability of peer-led mental health support intervention delivered by a mobile app, we conducted focus groups with participants of the study. The reporting of methods and findings adheres to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1) [34].

REACHOUT Intervention Description

Described in detail elsewhere, the REACHOUT pilot investigated the impact of a mobile app that delivered mental health support to adults with T1D living in Interior British Columbia over a period of 6 months. REACHOUT offered multiple support delivery modalities (one-on-one, group-based texting, and virtual face-to-face small group sessions that could be customized by recipients) [35]. Participants include individuals who receive support (recipients) and those who provide support (peer supporters). In this paper, the term “participants” will only be used when addressing both recipients and peer supporters. The eligibility criteria for recipients were as follows: (1) be diagnosed with T1D, (2) be at least 18 years or older, (3) speak English, (4) have access to the internet and/or a smartphone, (5) live in the interior region of British Columbia, and (6) have a mean subscale score of ≥2 on the type 1 Diabetes Distress Scale [36]. Peer supporters had similar requirements with the exceptions of criteria 5 and 6. They also had to be willing to complete a 6-hour training program. Training components and competency evaluation are published elsewhere [37]. It should be noted that if asked a medical question by recipients, peer supporters were instructed to refrain from answering and defer to the diabetes nurse educator.

The REACHOUT app offered multiple support delivery modalities including one-on-one support provided by a recipient-selected peer supporter, group texting support via the 24/7 chat room, and small group face-to-face support via video huddles and happy hours. Recipients were encouraged to use any or all modalities as frequently as desired. Peer supporters were invited to attend virtual wellness sessions to debrief their experiences as well as receive their own emotional support. Finally, the ongoing monitoring of group-based communication exchanges was performed by the research team, and fidelity assessments were conducted at 1, 3, and 5 months of the intervention with all participants.

Ethical Considerations

This qualitative descriptive study was approved by the University of British Columbia Behavioural Research Ethics Board (H20-00276). Prior to focus groups, participants provided e-informed consent using REDCap (Research Electronic Data Capture) electronic data capture tools hosted at the University of British Columbia [38,39]. To maintain privacy and confidentiality, recordings were anonymized to omit personal identifying information and stored securely. Only the study team could access study data. Upon completion, participants received a CAD $25 (approximately US $18) e-gift card.

Participant Recruitment and Sampling

Following the completion of the pilot trial REACHOUT, all those in recipient roles were contacted by a research assistant and invited to the postintervention focus groups to share their experience with the REACHOUT program and app and suggestions for improvement. Only peer supporters who had been paired with recipients were invited to join the

XSL•FO
RenderX

postintervention focus groups. Those who provided consent were interviewed.

Data Collection and Analysis

Focus groups were conducted online using Zoom; video and audio were recorded and later transcribed. Led by a female researcher (TST), focus groups were stratified into recipient versus peer supporter-only membership with approximately 6 individuals per group. The interview guide (Multimedia Appendix 1) used open-ended questions and prompts to elicit discussion around their experience in the program, peer support interactions, and app usage. Follow-up questions were posed if clarification or explanation was needed.

Recordings were transcribed verbatim and anonymized with participant roles (recipient or peer supporter) identified to capture perspectives from both groups. Transcripts were analyzed using NVivo V.14 software package [40]. Guided by an interpretivist research paradigm, which centers around subjective experiences [41], we selected an inductive thematic approach to support the possible variation of participant perceptions. Following Braun and Clarke's 6 phases of thematic analysis, 1 coder (DL) participated in transcribing the data and another coder (PJ) who had no involvement in the interview guide development, interviews, and transcription familiarized themselves with the transcripts [42]. Both coders discussed initial ideas before independently performing open coding. The coders discussed the findings after every round of coding to enhance reflexivity and iteratively refine a unified codebook. Independently coded transcripts were combined, and codes were sorted and combined to form themes and subthemes. Themes and subthemes were reviewed and refined with clear definitions and names. Findings and any discrepancies were discussed with the principal investigator (TST) and another coauthor member (DS) who was not involved in the interview guide creation and interviews. Moreover, this was a recursive process where analysis phases moved back and forth as needed [42].

Positionality Statement

Our multidisciplinary team comprises cisgender, heterosexual women from East Asian, South Asian, and European settler backgrounds. TST has over 25 years of experience working in peer support, and her research focuses on developing models to improve mental health outcomes in high-risk and medically underserved communities. DS has over 25 years of research working in diabetes self-management at the community and provider level. FSC has over 20 years of experience working on topics related to stress, social support, and social connection and contributes a behavioral science perspective. DL and PJ are early-career researchers with master's and medical graduate training. All authors are living in urban centers and are cognizant of their own privileges and practice reflexivity to ensure that priorities of the diabetes community are represented throughout the research process.

Results

Description of Sample

In total, 32 study participants (17 recipients and 15 peer supporters) who completed the REACHOUT intervention were recruited and interviewed from August to October 2022. The characteristics between focus group participants compared to nonrespondents in the pilot study population are noted in Multimedia Appendix 2. There were 9 focus groups lasting 60‑90 minutes, 4 recipient-only groups, and 5 peer supporter-only groups. As summarized in Table 1, participants were predominantly women and Caucasian, with a mean age of 48 (SD 16.3; range 23‑76) years and an average of 24 (SD 18.1; range 0‑65) years living with diabetes. Most participants received postsecondary education and had a household income greater than CAD $70,000 (approximately US $50,505).

XSL•FO
RenderX

Table . Interviewed recipients' and peer supporters' baseline characteristics.

	Total focus group participants (n=32)	Recipients (n=17)	Peer supporters (n=15)
Age (y), mean (SD)	48 (16.3)	48 (16.6)	50 (16.4)
Diabetes duration (y), mean (SD)	24 (18.1)	25 (18.5)	23 (18.2)
Women, n (%)	26 (81)	15 (88)	11 (73)
Marital status, n (%)			
Never married	9 (28)	6 (35)	3 (20)
Married or living with a partner	20 (63)	10 (59)	10 (67)
Separated or divorced or Widow	3 (9)	1 (6)	2 (13)
Ethnicity, n (%)			
Aboriginal	1 (3)	1 (6)	0 (0)
Aboriginal/Caucasian	1 (3)	1 (6)	0 (0)
East Asian (Chinese, Korean, Japanese)	1 (3)	0 (0)	1 (7)
Caucasian	29 (91)	15 (88)	14 (93)
Education, n (%)			
High school graduate (or equivalent)	3 (9)	3 (18)	0 (0)
Some college or technical school	7 (22)	4 (24)	3 (20)
College graduate	10 (31)	3 (18)	7 (47)
Graduate degree	12 (38)	7 (41)	5 (33)
Pretax household income (CAD $), n (%)			
<70,000 (approximately US $50,505)	10 (31)	7 (41)	3 (20)
>70,000 (approximately US $50,505)	17 (53)	5 (29)	12 (80)
Declined to answer	5 (16)	5 (29)	0 (0)
Employment, n (%)			
Full-time job	12 (38)	6 (35)	6 (40)
Part-time job	6 (19)	5 (29)	1 (7)
Retired	6 (19)	2 (12)	4 (27)
Other	7 (22)	4 (24)	3 (20)
Declined to answer	1 (3)	0 (0)	1 (7)

Themes

Four overarching themes were identified and related to participants' experiences in the peer support intervention and on their user experience with the mobile app delivery (Table 2).

XSL•FO
RenderX

Table . Four major themes were identified by thematic analysis with subthemes that capture similarities and differences within and across recipient and peer supporter groups.

Theme	Recipient	Peer supporter	Both group
Need for a sense of community and belonging	—[a]	—	• Giving and receiving unconditional support • Reducing isolation in rural communities • Learning from real-life experiences of T1D peers
Factors to enhance the recipient-peer supporter experience	• Ability to select a peer supporter	—	• Modality and frequency of communication
Key aspects of the peer supporter experience	—	• Supporting peer supporters in their role • Benefits of being a peer supporter • Challenges of being a peer supporter	—
Importance of personalizing the user experience while using the REACHOUT mobile app	—	—	• Varied preferences in peer support • Adapting the mobile app to fit user expectations

[a]Not applicable.

Theme 1: Need for a Sense of Community and Belonging

For recipients and peer supporters, REACHOUT created a safe environment to build and strengthen connections with other adults who shared the lived experience of T1D. This sense of belonging and community spirit manifested in different ways.

Subtheme A: Giving and Receiving Unconditional Support

The intervention created a space to express concerns without fear of judgment or rejection. Participants who had felt completely alone in the past finally found their "tribe"—a community that experienced and understood the same fears, frustrations, and emotional burdens of T1D.

> *The whole thing has been just so rewarding and I think it's kind of brought me out a little bit too. Like being able to be who I am and not be judged it's like – it's just this community. Being able to kind of hop into the chat and say, "Oh yeah this is what happened to me" or you know, just that common sharing. It's been huge.* [Peer supporter 5-2]

Initially, some participants were hesitant to engage in group activities such as face-to-face virtual sessions because the possibility of meeting peers who were managing their diabetes "perfectly" could trigger feelings of inadequacy or resentment. However, once the intervention started, they realized others were willing to be vulnerable. For example, when some participants disclosed perceived self-management failures in the 24/7 chat room, they were met with empathy and validation. After this precedent was established, others felt safe to reveal moments of insecurity and self-blame.

> *It was really nice to know when you're like, "I'm doing everything possible to keep my blood sugar stable right now and for the life of me they're on the higher side. I don't know why." But knowing other people are like, "Yeah, isn't that frustrating," like they get it because they live it. It's not like your [endocrinologist], it's nice to hear it from somebody who lives it, I don't feel so alone in the world.* [Recipient 6-3]

Subtheme B: Reducing Isolation in Rural Communities

Coming from rural and small communities across Interior British Columbia, many recipients and peer supporters had never encountered another T1D adult in their local community. This sense of loneliness was particularly pronounced for individuals diagnosed late in life (eg, 45 years and older).

> *It seems like we grew up in a smaller town, and there wasn't anybody that had diabetes that I knew, and then going through the other parts of my life, I didn't have really anybody to talk to.* [Recipient 4-2]

Although REACHOUT was a virtual intervention, participants were comforted knowing that peers resided in nearby towns. When browsing through the peer supporter library, participants were able to identify the general location where each peer supporter lived and, therefore, felt reassured that face-to-face support was accessible if needed. As part of the REACHOUT community, participants were not left to cope with the struggles of T1D on their own.

> *I thought it was really nice to connect with people, maybe not totally in my community. But certainly, there have been a great number of people within an hour's drive that's connected with and there's just something about that to know that you're not alone in your little portion of the world.* [Recipient 6-2]

XSL•FO
RenderX

Subtheme C: Learning From Real-Life Experiences of Peers With T1D

With REACHOUT, participants had direct access to the most reliable and high-quality T1D information including "real-world" experiences from adults who used insulin, insulin pumps, and continuous glucose monitors daily. The mobile app offered different mechanisms to obtain the knowledge needed. For instance, in the 24/7 chat room, participants posted updates regarding changes to health insurance coverage or, during the COVID-19 period, shortages in various diabetes supplies. This platform was also a place to pose questions and elicit differing perspectives from both recipients and peer supporters. For example, participants who were considering transitioning to a different insulin pump or continuous glucose monitoring device could hear opinions from peers from diverse lifestyles and backgrounds.

> *It was cool to hear firsthand information from somebody's experience, say about the Omnipod or the Medtronic or Dexcom or whatever. I think that's invaluable, rather than just going to a doctor or endocrinologist and just a medical professional, which is still really good information but to get the user's perspective on something is kind of for sure.* [Peer supporter 7-1]

Notably, how participants preferred to learn varied. Those who were not comfortable posting messages or disclosing personal experiences still enjoyed reading the discussion threads and exchanges in the 24/7 chat room. Many participants routinely checked the app to read the most recent conversation and updates. While not directly participating, participants who passively monitored the exchange of dialogue derived substantial benefits.

> *In my journey over the years with diabetes, I just felt so alone, so this app has been – just knowing it's there has been huge. I'm kind of a classic introvert – I don't really go on and participate actively on it, but I do on in and I read the conversations and just I love it. Please don't underestimate power of that because it's really been a big thing for me.* [Peer supporter 7-2]

Theme 2: Factors to Enhance the Recipient-Peer Supporter Experience

Factors related to one's experience with REACHOUT were largely dependent on the quality of the recipient-peer supporter relationship. Many found their peer supporter extremely helpful and valued their time, but the strength of their relationship was influenced by various contributing factors.

Subtheme A: Ability to Select a Peer Supporter

Recipients felt empowered by the opportunity to choose their peer supporter. Some sought identical counterparts, while others envisioned their peers as potential mentors. The criteria that each recipient used to choose their peer supporter were unique and personal. The main factors included diabetes management system, duration of diabetes, shared activities, life stage, and age.

According to some recipients, diabetes and management-related factors weighed heavily into the selection process. For example, some recipients were seeking a peer supporter who had been living with diabetes for as long, if not longer, than themselves. Others felt a greater kinship with peer supporters using the same continuous glucose monitoring or insulin pump.

> *I looked at not necessarily insulin type, but just device that they might be using. And for me, the Dexcom was new so I wanted somebody who knew and used the Dexcom. So that was some of my criteria when I started to go through the list. I don't need to read the other fifteen that don't use a Dexcom, that was a clear priority for me.* [Recipient 6-3]

Lifestyle factors also factored in prominently when selecting a peer supporter. For instance, recipients who enjoyed exercising or engaging in outdoor sports preferred an equally active peer supporter. Having shared hobbies enhanced the quality of recipient-peer supporter relationships and extended conversations beyond the boundaries of diabetes. In contrast, in the absence of similar interests, some recipients found it difficult to establish meaningful and sustained rapport with their peer supporters.

> *Device for me wasn't as important. Cause I've been on both injections and pump. So for me, mostly activities and hobbies. And someone that liked to travel as well, cause I always find that quite daunting but I want to do more of that so yeah. I found a good person for that.* [Recipient 8-4]

The stage of life was equally important. For example, young mothers gravitated toward selecting peer supporters who were also raising children. As expected, navigating both diabetes and parenthood created strong connections. Similarly, older recipients who were retired understood the priorities and pace of others who were also no longer in the workforce.

> *I picked someone who was in a similar life stage as me, cause I've had diabetes for 30 years I don't really need advice on how to treat my diabetes. For me, it was much more the mental health connection and then transition to this new part of my life of being a mom. Because stuff would come up and I'd be like oh, my gosh, how do you deal with this? How do you prioritize a crying baby verses a low? So that for me was great.* [Recipient 9-1]

Age and/or length of diabetes experience emerged as critical factors in the selection process. Some recipients intentionally chose older peer supporters who had a lifelong journey with diabetes as they envisioned having a mentor who could provide insight on what challenges to expect over time. Rarely did recipients choose peer supporters who were much younger than themselves.

> *Someone [who] was male, and older than me. So I can relate to what they're going through, and someone who has had diabetes for longer than I have. So it's quite focused of what I was looking for. I was able to be paired up with someone who was in my*

XSL•FO
RenderX

position, but a couple years down the road. [Recipient 8-3]

Subtheme B: Modality and Frequency of Communication

Video conferencing was the most preferred modality, as it allowed for the 2 parties to observe facial expressions and body language. Different communication methods were utilized for different functions. Direct messaging, texting, and emails were ideal for quick communication such as check-ins and meeting coordination. If both parties were amenable to investing greater effort and commitment, more substantial conversations took place through video conferencing or phone calls.

Consistency formed the foundation of a strong recipient-peer supporter relationship. Initially, weekly communication was needed to establish and build rapport. However, as the relationship matured, for some, the frequency of contact slowed down as people had other competing life demands such as full-time jobs or home responsibilities. Mid-intervention, many acknowledged that the ideal schedule was contact once every 2 weeks.

> *I liked that it was once a week in the beginning. I think it gave you a lot of opportunity to get to know each other, tell each other your diagnosis story and then from there on. I think I did realize with my peer supporter when we started, when we were meeting every week that we almost were running out of things to update each other on or talk about. And then every two weeks was really great and then we had some things to share over the last two weeks.* [Recipient 8-1]

Theme 3: Key Aspects of the Peer Supporter Experience

The cornerstone of a peer-led intervention is the peer supporters who deliver mental health support. Although the goal of REACHOUT was to provide support to recipients, the sustained quality of the 6-month intervention provided opportunities for peer supporters to be nurtured as well.

Subtheme A: Supporting Peer Supporters in Their Role

To function effectively in their role, peer supporters underwent a 6-hour training. According to peer supporters, the most instrumental training activity was “role-plays.” Not only did role-plays allow trainees to practice newly developed skills, but these simulated scenarios helped build their self-confidence and preparedness.

During the intervention, peer supporters appreciated having a workbook with structured activities to lead their recipients through. These activities served as a valuable foundation for conversations that would not occur organically—for example, identifying personal values and exploring sources of diabetes distress.

Furthermore, peer supporters benefited from attending wellness sessions hosted by the research team. Wellness sessions were Zoom-based and provided the opportunity for peer supporters to share stories, voice concerns, and pose questions to one another. Moreover, these discussions fostered camaraderie among peer supporters while navigating inherent challenges in their support roles.

> *I think every [Wellness] session – I found important, because there’s always something new that you can take away. And then, if there’s a question that I have, [I] can actually ask during those sessions. “Okay, you know. Great. I’m on the right track,” you know as well and then, “I’m following what I supposed to be following and doing what I’m supposed to be doing with the peers.”* [Peer supporter 1-1]

Subtheme B: Benefits of Being a Peer Supporter

Peer supporters derived deep satisfaction and intrinsic reward from their role, finding genuine fulfillment from providing mental health support to other adults with T1D. Through acts of altruism and compassion for the T1D community, they experienced satisfaction knowing that their contribution added meaning and value to the lives of their recipient.

Many peer supporters realized that their relationship with their matched recipient was mutually beneficial. Not only did peer supporters deliver emotional support, but recipients also shared their knowledge, coping strategies, and perspectives. Additionally, many peer supporters discovered a renewed connection with their own diabetes journey and engaged in self-reflection and self-development.

> *[My recipient] was fairly newly diagnosed, within the last year, and it’s been 11 years for me. I benefited a lot from talking with her. It kind of re-engaged me in diabetes. I think I realized I’ve been coasting, and I needed to kind of re-engage, and I think that was really important for me.* [Peer supporter 2-2]

Subtheme C: Challenges of Being a Peer Supporter

Not all peer supporters had recipients who reciprocated with the same level of enthusiasm. Rather than feeling rejected if their recipient did not respond immediately, some peer supporters did not take it personally. Moreover, peer supporters found it challenging to sustain consistent communication with their recipients, especially in the last half of the intervention. Peer supporters tried to understand their recipient’s perspectives by acknowledging the demands of personal and professional lives.

> *I found sending a text- something, I felt like I was chasing her. And I would think, “Oh maybe she doesn’t want to talk to me anymore,” “Maybe she’s had enough,” or, “Maybe I’m doing something wrong,” but it wasn’t anything like that at all. It was just she was busy; she has a job and family.* [Peer supporter 5-3]

Some peer supporters struggled to deepen their conversations when recipients appeared to be reluctant to broach more sensitive topics. At times, peer supporters adhered to surface-level conversations so as to not “over-step.” As such, peer supporters suggested having more guidance on how to navigate boundaries and tips for gauging the depth recipients seek from relationships.

XSL•FO
RenderX

> *I didn't bring up the underlying issues as much as I would have expected, perhaps because I wasn't quite versed in how to bring those up. I didn't know if it was appropriate for me to kind of prod a little bit. […] I felt a little bit at a loss of how to bring up like these big concepts, psychological issues and things like that. There was definitely stuff going on, but it was hard for me to get them to speak about some of those things.* [Peer supporter 2-1]

Conversely, some peer supporters encountered recipients who openly shared their feelings and concerns, which posed a different challenge as it triggered feelings of worry and inadequacy. Peer supporters were seeking greater instructions on how to navigate these emotionally charged conversations. Two potential solutions suggested were (1) establishing clear guidelines on how to respond to questions requiring escalation to a health professional and (2) providing a set of prepared questions to ask when these situations arose.

> *I'm not gonna lie, I was a little bit stressed if this person was really in distress, because I don't know if I was like, "Jeez, like I don't know if I can be the guy that's going to help this person." But I was pretty fortunate, [my recipient] just wants someone to talk to, basically, which worked out well for me.* [Peer supporter 7-1]

Theme 4: Importance of Personalizing the User Experience While Using the REACHOUT Mobile App

Participants (recipients and peer supporters) had four ways to engage with others on the REACHOUT mobile app: (1) direct messaging, (2) 24/7 chat room, (3) virtual happy hours, and (4) virtual huddles.

Subtheme A: Varied Preferences in Peer Support

The 24/7 group chat room served as a central feature of the app with a significant amount of activity. Most participants referred to the 24/7 chat room to pose questions, share stories and updates, and initiate discussions. The high level of engagement led many participants to habitually check the chat to stay informed. For some, monitoring the 24/7 chat room was a part of their daily routine, as participants could obtain new information as well as be exposed to a diverse range of topics.

Alternatively, some found the continuous flow of information in the 24/7 chat room to be overwhelming. Specifically, it was burdensome to sift through a high number of messages to find discussions of personal relevance. For example, while the majority of participants discussed insulin pumps, it alienated the few individuals who used multiple daily injections. In extreme cases, some participants deactivated the notifications setting for the 24/7 chat room.

> *Like it was overwhelming right at the beginning [from the 24/7 chat room], and so I turned off the notifications but then I got it out of the habit of checking, so I missed a whole bunch of stuff, me and my mentor were communicating through text, so I didn't really have to worry about going back into the app.* [Recipient 9-1]

Virtual huddles and virtual happy hours were 2 additional support delivery mechanisms offered. The former was a larger-group interactive webinar led by peer supporters and/or professionals and required fewer social demands or active participation. The latter involved a smaller, intimate group discussion led by a peer supporter and fostered open and relaxed conversations beyond their one-on-one peer support relationship. These 2 support modalities cater to diverse personality types and needs.

Subtheme B: Adapting the Mobile App to Fit User Expectations

App usability issues centered largely around the lack of logical structure and flow of exchanges within the 24/7 chat room. Because participants had the option of responding within a thread or creating a new thread, conversations often seemed disjointed. As a result, many suggested creating more topic-focused discussion boards as "exit ramps" from the 24/7 chat room, allowing participants to select personally relevant information in a structured way. Participants also suggested a keyword search feature. This element would streamline the process of finding specific information without the need to scroll through recent posts. To increase accessibility for people with different reading abilities, participants suggested that the app be available on bigger devices such as tablets or computers.

Finally, the mobile app experienced various bugs. For research purposes, this app was launched on a testing platform that required participants to log in with their credentials every 3 months. This issue led to widespread frustration and confusion among participants who lost access unexpectedly. Additionally, there were bugs in the video feature, which made it difficult for participants to connect unless they used platforms outside of the mobile app (eg, Zoom, Facetime). Future improvements to fix these bugs would ensure a smoother and more reliable user experience.

> *I guess I went to log on the other day I wasn't sure when it ended, and I was quite sad when I didn't have access anymore, to go on and read the stuff I was used to reading each day so that was kind of, that was nice. Well, it wasn't nice that I couldn't get on but it was nice, yeah.* [Recipient 4-2]

Discussion

Principal Findings

This study explored recipients' and peer supporters' experiences with and perspectives on REACHOUT, a peer-led mental health support intervention for adults with T1D living in rural and remote regions of British Columbia. Our results identified four major themes: (1) Need for a sense of community and belonging, (2) Factors to enhance the recipient-peer supporter experience, (3) Key aspects of the peer supporter experience, and (4) Importance of personalizing the user experience while using the REACHOUT mobile app.

Comparison to Prior Work

Consistent with our findings, the need for community and belonging, especially for geographically marginalized

XSL•FO
RenderX

individuals, has also been reported in the literature. For instance, a systematic review of 12 qualitative studies on health care access for rural patients with chronic diseases found that a sense of group connection in rural areas mitigates feelings of vulnerability [43]. Similarly, Joensen et al [44] noted that while a feeling of inclusion contributes to health promotion, it is often lacking in daily life for individuals with T1D. Thus, a mobile app such as REACHOUT is especially valuable in addressing these gaps in remote and underserved communities.

With REACHOUT, recipients had the agency to choose a peer supporter based on personally relevant factors. This choice-based model deserves consideration, as it may optimize the recipient-peer supporter match [12,35]. Our data also suggest that successful pairs often referred to one another as "friends," which supports the idea that effective emotional support is built upon friendship and trust [17,22]. To enhance participant satisfaction, future peer support studies should adopt this recipient-driven matching process as recipients are in the best position to understand their own unique support needs.

While the one-on-one and group support delivery mechanisms address different support needs, many recipients expressed greater value for the former. The advantages of personalized individual relationships address the limitations inherent in group settings. For example, in an intervention of peer support meetings for adults with T1D focusing on insulin pumps, dissatisfied participants reported a lack of relevance in the discussion topics, hindering their ability to speak about topics that mattered to them [45]. Incorporating modalities that allow recipients to seek both group-based and one-on-one peer support within the same intervention promotes greater support customization for each user. Subsequent mental health support models should prioritize flexible delivery options that balance individualized support with opportunities for group engagement.

As participation in group activities within the mobile app was optional, we observed varying levels of engagement. Passive participation, characterized by viewing (vs posting) 24/7 chat room exchanges, was the most common. Participants engaged in "lurking" behavior, which involved routinely checking the chat room, gleaning value in reading anecdotes and being exposed to new topics related to T1D. "Lurking" was also observed in an online community–based peer support forum for in-hospital patients. This study found that 7 of 30 participants opted not to post yet still experienced a positive impact on emotional well-being [46]. Additionally, Tang et al [47] found that adults with T1D who passively engaged with the digital support platform (ie, 'lurkers') reported greater reductions in stigma-related distress compared to active posters. These findings highlight the role of passive engagement in mental health interventions as a strategy for mitigating "social risks" [47,48]. An in-depth examination of the mental health benefits associated with passive participation on digital platforms is warranted.

While not anticipated by peer supporters, the flow of support with recipients was bidirectional. However, the content of the "give and take" exchange likely encompassed a range of topics not necessarily diabetes-specific. Nonetheless, this opportunity for mutual sharing was also cited in a systematic review of qualitative peer support studies for chronic diseases [49]. Recognizing this reciprocity as an unintentional intervention, peer support studies should routinely assess changes in outcomes for both recipients and peer supporters. Clearly, peer support fosters emotional well-being for both parties.

Ensuring ongoing support for peer supporters beyond the initial training phase is essential for peer supporter effectiveness and well-being. Our intervention addressed this need by offering peer wellness sessions, a space for peer supporters to share successes and challenges. Not surprisingly, emotional investment leading to exhaustion can harm the mental health of peer supporters [50]. Thus, having an environment to express frustrations in real time such as how to deal with nonresponsive recipients or navigate emotionally charged conversations could potentially prevent burnout or dissatisfaction. Therefore, implementing regular communication or check-ins could enhance peer supporters' experience and overall intervention effectiveness.

Limitations

First, this study only recruited matched peer supporters (vs unmatched peer supporters). Perspectives from unmatched peer supporters were not captured. Future studies should consider interviewing those peer supporters who did not participate in the one-on-one support component but had access to other support delivery features. Second, this sample was self-selected and possibly more engaged and enthusiastic than other participants. This may limit the representativeness of the original REACHOUT cohort. While we compared the characteristics of the consenting and nonconsenting sample, future studies should ensure representation across different levels of engagement. Third, the socioeconomic background for participants was relatively high. Because we did not overrecruit for individuals with lower levels of income or education, the diversity of experiences captured may be skewed. Finally, the study targeted the rural and remote communities of Interior British Columbia; therefore, the results may not be generalizable to other geographically marginalized populations in BC or Canada.

Conclusions

Peer support is increasingly recognized as a critical component for mental health interventions in T1D. While research has focused largely on recipients of support, our study also considered perspectives of individuals delivering support, providing a holistic view. More importantly, it is the recipient-peer supporter dynamic that most likely drives the success of the implementation of the REACHOUT program and impacts mental health outcomes. Only by understanding the experiences of both parties can we refine our interventions to provide the optimal mental health support model.

XSL•FO
RenderX

Acknowledgments

We would like to acknowledge Interior Health, Diabetes Education Centres, health care professionals, and research team who supported the recruitment of recipients and peer supporters. This study was supported by Breakthrough T1D International 2-SRA-2020-986-S-B. The authors declare that there was no use of generative artificial intelligence in the writing, editing, or development of this paper. The focus group recordings were transcribed using the Zoom transcription service and the authors reviewed all transcripts to ensure accuracy and deidentified personal information.

Funding

TST's program of research is funded by Michael Smith Health Research BC/Breakthrough T1D Canada Health Professional-Investigator Award #HPI-2021-2359.

Data Availability

The dataset generated and analyzed during this study is not publicly available to ensure participants' privacy. For questions about the dataset, contact the corresponding author.

Authors' Contributions

TST contributed to funding acquisition, study conceptualization, design, implementation, supervision, manuscript review, editing, and revision and is the study guarantor. DL contributed to the study implementation, data collection, analysis, interpretation, original manuscript preparation, editing, and revision. PJ contributed to the data analysis and interpretation. DS contributed to the data interpretation, manuscript review, and editing. FSC contributed to data interpretation, manuscript review, and editing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide created by the principal investigator and research team to guide focus group discussions.
[DOCX File, 20 KB - diabetes_v11i1e72779_app1.docx]

Multimedia Appendix 2
Interviewed participants' baseline characteristics compared to nonrespondents from the larger pilot participant population. Mann-Whitney *U* tests were applied for continuous variables and the Fisher exact test for categorical variables.
[DOCX File, 18 KB - diabetes_v11i1e72779_app2.docx]

Checklist 1
COREQ: 32-item checklist.
[DOCX File, 27 KB - diabetes_v11i1e72779_app3.docx]

References

1. Bennett PN, St Clair Russell J, Atwal J, Brown L, Schiller B. Patient-to-patient peer mentor support in dialysis: improving the patient experience. Semin Dial 2018 Sep;31(5):455-461. [doi: 10.1111/sdi.12703] [Medline: 29722062]
2. Kiemen A, Czornik M, Weis J. How effective is peer-to-peer support in cancer patients and survivors? A systematic review. J Cancer Res Clin Oncol 2023 Sep;149(11):9461-9485. [doi: 10.1007/s00432-023-04753-8] [Medline: 37120782]
3. Longley RM, Harnedy LE, Ghanime PM, et al. Peer support interventions in patients with kidney failure: a systematic review. J Psychosom Res 2023 Aug;171:111379. [doi: 10.1016/j.jpsychores.2023.111379] [Medline: 37270909]
4. Thompson DM, Booth L, Moore D, Mathers J. Peer support for people with chronic conditions: a systematic review of reviews. BMC Health Serv Res 2022 Mar 31;22(1):427. [doi: 10.1186/s12913-022-07816-7] [Medline: 35361215]
5. Azmiardi A, Murti B, Febrinasari RP, Tamtomo DG. The effect of peer support in diabetes self-management education on glycemic control in patients with type 2 diabetes: a systematic review and meta-analysis. Epidemiol Health 2021;43:e2021090. [doi: 10.4178/epih.e2021090] [Medline: 34696569]
6. Chen C, Zhou Y, Xu JY, Song HY, Yin XW, Gu ZJ. Effect of peer support interventions in patients with type 2 diabetes: a systematic review. Patient Educ Couns 2024 May;122(108172). [doi: 10.1016/j.pec.2024.108172] [Medline: 38320444]
7. Krishnamoorthy Y, Sakthivel M, Sarveswaran G, Eliyas SK. Effectiveness of peer led intervention in improvement of clinical outcomes among diabetes mellitus and hypertension patients—a systematic review and meta-analysis. Prim Care Diabetes 2019 Apr;13(2):158-169. [doi: 10.1016/j.pcd.2018.11.007] [Medline: 30528217]

XSL•FO
RenderX

8. Lu S, Leduc N, Moullec G. Type 2 diabetes peer support interventions as a complement to primary care settings in high-income nations: a scoping review. Patient Educ Couns 2022 Nov;105(11):3267-3278. [doi: 10.1016/j.pec.2022.08.010] [Medline: 36038395]
9. Grant V, Litchfield I. Acceptability of community health worker and peer supported interventions for ethnic minorities with type 2 diabetes: a qualitative systematic review. Front Clin Diabetes Healthc 2024;5:1306199. [doi: 10.3389/fcdhc.2024.1306199] [Medline: 38836261]
10. Okoro FO, Veri S, Davis V. Culturally appropriate peer-led behavior support program for African Americans with type 2 diabetes. Front Public Health 2018;6:340. [doi: 10.3389/fpubh.2018.00340] [Medline: 30533408]
11. Shiyanbola OO, Maurer M, Schwerer L, et al. A culturally tailored diabetes self-management intervention incorporating race-congruent peer support to address beliefs, medication adherence and diabetes control in African Americans: a pilot feasibility study. Patient Prefer Adherence 2022;16:2893-2912. [doi: 10.2147/PPA.S384974] [Medline: 36317056]
12. Afshar R, Askari AS, Sidhu R, et al. Out of the mouths of peer leaders: perspectives on how to improve a telephone-based peer support intervention in type 2 diabetes. Diabet Med 2022 Sep;39(9):e14853. [doi: 10.1111/dme.14853] [Medline: 35437815]
13. Allen NA, Litchman ML, Greenwood DA, Ng A, Sanchez-Birkhead A. Hispanic diabetes peer facilitator training needs and connections to Hispanics with type 2 diabetes. Hisp Health Care Int 2023 Dec;21(4):221-227. [doi: 10.1177/15404153231175998] [Medline: 37221843]
14. Ayala GX, Canale K, Ibarra L, et al. Facilitators of peer leader retention within a type 2 diabetes intervention trial for US Latinos/Hispanics of Mexican origin. Health Educ Res 2023 May 22;38(3):204-219. [doi: 10.1093/her/cyad006] [Medline: 36848036]
15. Goldman ML, Ghorob A, Eyre SL, Bodenheimer T. How do peer coaches improve diabetes care for low-income patients?: a qualitative analysis. Diabetes Educ 2013;39(6):800-810. [doi: 10.1177/0145721713505779] [Medline: 24168838]
16. Holman D, Simmons D, Ockenden N, Graffy JP. Learning from the experience of peer support facilitators and study nurses in diabetes peer support: a qualitative study. Chronic Illn 2021 Sep;17(3):269-282. [doi: 10.1177/1742395319873378] [Medline: 31495199]
17. Kowitt SD, Urlaub D, Guzman-Corrales L, et al. Emotional support for diabetes management: an international cross-cultural study. Diabetes Educ 2015 Jun;41(3):291-300. [doi: 10.1177/0145721715574729] [Medline: 25722064]
18. Tarfa A, Nordin J, Mott M, Maurer M, Shiyanbola O. A qualitative exploration of the experiences of peer leaders in an intervention to improve diabetes medication adherence in African Americans. BMC Public Health 2023 Jan 20;23(1):144. [doi: 10.1186/s12889-023-15059-2] [Medline: 36670394]
19. Wen MJ, Salihu EY, Yang C, Maurer M, Shiyanbola OO. Peer ambassador perspectives in a culturally tailored self-management intervention for African Americans with type 2 diabetes: a qualitative study. Pharmacy (Basel) 2024 May 8;12(3):75. [doi: 10.3390/pharmacy12030075] [Medline: 38804467]
20. Dale J, Caramlau I, Sturt J, Friede T, Walker R. Telephone peer-delivered intervention for diabetes motivation and support: the telecare exploratory RCT. Patient Educ Couns 2009 Apr;75(1):91-98. [doi: 10.1016/j.pec.2008.09.014] [Medline: 19013741]
21. Kim C, MacMillan F, Osuagwu UL, Simmons D. Peer support facilitator and peer perspectives of an inner-regional Australian diabetes prevention and self-management peer support program. Health Soc Care Community 2022 Nov;30(6):e4051-e4064. [doi: 10.1111/hsc.13798] [Medline: 35318758]
22. Lott BD, Dicks TN, Keddem S, Ganetsky VS, Shea JA, Long JA. Insights into veterans' perspectives on a peer support program for glycemic management. Diabetes Educ 2019 Dec;45(6):607-615. [doi: 10.1177/0145721719879417] [Medline: 31596174]
23. Paul G, Keogh K, D'Eath M, Smith SM. Implementing a peer-support intervention for people with type 2 diabetes: a qualitative study. Fam Pract 2013 Oct;30(5):593-603. [doi: 10.1093/fampra/cmt027] [Medline: 23836093]
24. Ashrafi S, Taylor D, Tang TS. Moving beyond "don't ask, don't tell": mental health needs of adults with type 1 diabetes in rural and remote regions of British Columbia. Diabet Med 2021 May;38(5):e14534. [doi: 10.1111/dme.14534] [Medline: 33524209]
25. Bujanda-Sainz de Murieta A, Escalada-Hernández P, San Martín-Rodríguez L, Soto-Ruiz N. Determinants of the adoption of an online community for people with diabetes: a qualitative study. Sci Rep 2025 Oct 22;15(1):36944. [doi: 10.1038/s41598-025-20960-4] [Medline: 41125712]
26. Malik FS, Lind C, Duncan S, Mitrovich C, Pascual M, Yi-Frazier JP. Augmenting traditional support groups for adolescents with type 1 diabetes using Instagram: mixed methods feasibility study. JMIR Diabetes 2021 Oct 21;6(4):e21405. [doi: 10.2196/21405] [Medline: 34673527]
27. Litchman ML, Oser TK, Hodgson L, et al. In-person and technology-mediated peer support in diabetes care: a systematic review of reviews and gap analysis. Diabetes Educ 2020 Jun;46(3):230-241. [doi: 10.1177/0145721720913275] [Medline: 32321370]
28. Berkanish P, Pan S, Viola A, Rademaker Q, Devine KA. Technology-based peer support interventions for adolescents with chronic illness: a systematic review. J Clin Psychol Med Settings 2022 Dec;29(4):911-942. [doi: 10.1007/s10880-022-09853-0] [Medline: 35147830]

XSL•FO
RenderX

29. Titoria R, Fung A, Tang TS, Amed S. Systematic review of technology-mediated peer support interventions in paediatric type 1 diabetes care. Diabet Med 2023 Oct;40(10):e15172. [doi: 10.1111/dme.15172] [Medline: 37428650]
30. Berg M, Linden K, Adolfsson A, Sparud Lundin C, Ranerup A. Web-based intervention for women with type 1 diabetes in pregnancy and early motherhood: critical analysis of adherence to technological elements and study design. J Med Internet Res 2018 May 2;20(5):e160. [doi: 10.2196/jmir.9665] [Medline: 29720365]
31. Linden K, Berg M, Adolfsson A, Sparud-Lundin C. Person-centred, web-based support in pregnancy and early motherhood for women with type 1 diabetes mellitus: a randomized controlled trial. Diabet Med 2018 Feb;35(2):232-241. [doi: 10.1111/dme.13552] [Medline: 29171071]
32. Liu Z, Wang C, Yang D, et al. High engagement in mobile peer support is associated with better glycemic control in type 1 diabetes: a real-world study. J Diabetes Investig 2022 Nov;13(11):1914-1924. [doi: 10.1111/jdi.13870] [Medline: 35708894]
33. Xie LF, Housni A, Nakhla M, et al. Adaptation of an adult web application for type 1 diabetes self-management to youth using the behavior change wheel to tailor the needs of health care transition: qualitative interview study. JMIR Diabetes 2023 Apr 26;8:e42564. [doi: 10.2196/42564] [Medline: 37121571]
34. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
35. Tang TS, Klein G, Görges M, et al. Evaluating a mental health support mobile app for adults with type 1 diabetes living in rural and remote communities: the REACHOUT pilot study. Diabet Med 2025 Mar;42(3):e15451. [doi: 10.1111/dme.15451] [Medline: 39538423]
36. Fisher L, Polonsky WH, Hessler DM, et al. Understanding the sources of diabetes distress in adults with type 1 diabetes. J Diabetes Complications 2015;29(4):572-577. [doi: 10.1016/j.jdiacomp.2015.01.012] [Medline: 25765489]
37. Tang TS, Yip AKW, Klein G, et al. Training peers to deliver mental health support to adults with type 1 diabetes using the REACHOUT mobile app. Diabet Med 2024 Mar;41(3):e15210. [doi: 10.1111/dme.15210] [Medline: 37634222]
38. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
39. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
40. NVivo – leading qualitative data analysis software. Lumivero. URL: https://lumivero.com/products/nvivo/ [accessed 2026-01-07]
41. Ryan G. Introduction to positivism, interpretivism and critical theory. Nurse Res 2018 Mar 16;25(4):14-20. [doi: 10.7748/nr.2018.e1466] [Medline: 29546962]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Brundisini F, Giacomini M, DeJean D, Vanstone M, Winsor S, Smith A. Chronic disease patients' experiences with accessing health care in rural and remote areas: a systematic review and qualitative meta-synthesis. Ont Health Technol Assess Ser 2013;13(15):1-33. [Medline: 24228078]
44. Joensen LE, Filges T, Willaing I. Patient perspectives on peer support for adults with type 1 diabetes: a need for diabetes-specific social capital. Patient Prefer Adherence 2016;10:1443-1451. [doi: 10.2147/PPA.S111696] [Medline: 27536076]
45. Joensen LE, Meldgaard Andersen M, Jensen S, Nørgaard K, Willaing I. The effect of peer support in adults with insulin pump-treated type 1 diabetes: a pilot study of a flexible and participatory intervention. Patient Prefer Adherence 2017;11:1879-1890. [doi: 10.2147/PPA.S142204] [Medline: 29138541]
46. Haldar S, Mishra SR, Kim Y, Hartzler A, Pollack AH, Pratt W. Use and impact of an online community for hospital patients. J Am Med Inform Assoc 2020 Apr 1;27(4):549-557. [doi: 10.1093/jamia/ocz212] [Medline: 31986197]
47. Tang TS, Seddigh S, Halbe E, Vesco AT. Testing 3 digital health platforms to improve mental health outcomes in adults with type 1 diabetes: a pilot trial. Can J Diabetes 2024 Feb;48(1):18-25. [doi: 10.1016/j.jcjd.2023.08.006] [Medline: 37625504]
48. Mazuro C, Rao N. Online discussion forums in higher education: is 'lurking' working? Int J cross-disciplinary Subj Educ 2011;2(2):364-371. [doi: 10.20533/ijcdse.2042.6364.2011.0051]
49. Embuldeniya G, Veinot P, Bell E, et al. The experience and impact of chronic disease peer support interventions: a qualitative synthesis. Patient Educ Couns 2013 Jul;92(1):3-12. [doi: 10.1016/j.pec.2013.02.002] [Medline: 23453850]
50. Kane L, Portman RM, Eberhardt J, et al. Peer supporters' mental health and emotional wellbeing needs: key factors and opportunities for co-produced training. Health Expect 2023 Dec;26(6):2387-2395. [doi: 10.1111/hex.13836] [Medline: 37565576]

XSL•FO
RenderX

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

Edited by A Stone; submitted 17.Feb.2025; peer-reviewed by D Tilden, L Fried, M Pulier; revised version received 19.Dec.2025; accepted 22.Dec.2025; published 21.Jan.2026.

Please cite as:
Lam D, Jhajj P, Sherifali D, Chen FS, Tang TS
Exploring the REACHOUT Mental Health Support App for Type 1 Diabetes From the Perspectives of Recipients and Providers of Peer Support: Qualitative Study
JMIR Diabetes 2026;11:e72779
URL: https://diabetes.jmir.org/2026/1/e72779
doi:10.2196/72779

© Debbie Lam, Poonamdeep Jhajj, Diana Sherifali, Frances S Chen, Tricia S Tang. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 21.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Implementing Diabetic Eye Care in Rural Primary Care Through Teleophthalmology and Pop-Up Clinics: Mixed Methods Study

Nisha G Arya[1], BA; Victoria Koltchine[1]; Talia C Gearinger[2], MPH; Sherianne Buehler[2], MSW; Reza Yousefi-Nooraie[3,4], PhD; Rajeev S Ramchandran[2,3,4], MD, MBA

[1]University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Rochester, NY, United States

[2]Flaum Eye Institute, University of Rochester Medical Center, Rochester, NY, United States

[3]Environmental Medicine and Public Health Sciences, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States

[4]Center for Community Health and Prevention, University of Rochester, Rochester, NY, United States

Corresponding Author:
Nisha G Arya, BA
University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Rochester, NY, United States

Abstract

Background: Diabetic retinopathy is the leading cause of blindness in US adults. Fewer than half of patients with diabetes in rural areas complete the recommended screening that could prevent blindness. Store-and-forward teleophthalmology and mobile pop-up eye clinics have each been shown to improve screening access, but studies examining the implementation of both modalities within the same rural primary care system are lacking.

Objective: This study aimed to identify multilevel facilitators and barriers to implementing teleophthalmology and pop-up eye clinics for diabetic eye care in 2 rural primary care clinics in upstate New York.

Methods: We conducted a mixed methods implementation study with prospective semistructured interviews and retrospective, descriptive electronic health record data at 2 university-affiliated primary care clinics classified as “isolated rural,” guided by the Consolidated Framework for Implementation Research (CFIR). Electronic health record data from 885 adults with diabetes were used to characterize patient demographics and eye care use. Patients were categorized into 3 groups: eye exam with an eye doctor (n=586), on-site primary care–based eye exam (n=91; teleophthalmology n=39 and pop-up clinic n=52), or no recent eye exam (n=208). Semistructured interviews were conducted with 20 patients (n=5 per subgroup) and 14 staff members (primary care clinicians, optometrists, and administrative personnel), purposively sampled to capture diverse perspectives across eye care pathways and roles. Interviews were coded using CFIR domains and analyzed thematically until no new themes emerged.

Results: Patients in the on-site primary care eye exam group were the most socially disadvantaged, with the lowest proportions reporting no transportation needs (42/91, 46.2%), no housing instability (28/91, 30.8%), and no food insecurity (36/91, 39.6%). Over 18 months, 58.4% (52/89) of scheduled pop-up clinic appointments were completed, and 92.3% (36/39) of teleophthalmology images were gradable. Pop-up clinics detected higher rates of diabetic retinopathy (11/52, 21.2% vs 3/39, 7.7%), cataract (29/52, 55.8% vs 3/39, 7.7%), and reduced visual acuity (23/52, 44.2% vs 11/39, 28.2%) than teleophthalmology. Qualitative analysis revealed that patients viewed both modalities as a convenient “one-stop shop” and valued trust in primary care staff. Staff identified complementary barriers: training and workflow complexity for teleophthalmology, and underutilization and scheduling challenges for pop-up clinics. Implementation champions across roles proposed an integrated workflow in which same-day teleophthalmology becomes a default component of annual diabetes visits, with targeted referral to pop-up clinics for patients with abnormal findings.

Conclusions: Primary care–based eye programs in rural settings preferentially reached socially vulnerable patients with diabetes and detected substantial unmet eye pathology. However, operating teleophthalmology and pop-up clinics as separate programs led to inefficiencies in both programs. A stakeholder-driven integrated model embedding teleophthalmology into routine diabetes care, with targeted pop-up clinic referrals, may improve the reach, efficiency, and sustainability of diabetic eye screening in rural populations.

(JMIR Diabetes 2026;11:e106290) doi:10.2196/106290

KEYWORDS

teleophthalmology; mobile health units; diabetic retinopathy; implementation science; mixed methods research; Consolidated Framework for Implementation Research

XSL•FO
RenderX

Introduction

Diabetic retinopathy is the leading cause of blindness in working-age US adults, with higher prevalence in rural than urban populations [1,2]. Although annual dilated eye examinations can prevent blindness in more than 95% of patients with diabetes, fewer than half of patients in underserved rural areas complete this recommended screening [2-4]. Rural communities face compounding barriers to eye care access, including severe workforce shortages: nearly 63% of US counties have no practicing ophthalmologist, and rural ophthalmologist density is roughly one-tenth that of metropolitan areas [1]. These shortages are projected to worsen, with ophthalmology workforce adequacy in nonmetropolitan areas forecast to fall to 29% by 2035, compared with 77% in metropolitan areas [4]. Beyond workforce gaps, rural patients face geographic isolation, limited transportation, lower socioeconomic status, and higher rates of uninsurance, creating eye care deserts in which patients with diabetes are at elevated risk for undetected and untreated diabetic retinopathy [5-11].

Two delivery models have been developed to bring diabetic eye screening to patients in primary care settings. Store-and-forward teleophthalmology uses nonmydriatic retinal photography obtained by trained primary care staff, with the images transmitted to a remote eye care specialist for interpretation. Previous programs have demonstrated effectiveness in improving diabetic retinopathy detection and screening rates [5-10]. At the University of Rochester, teleophthalmology initially increased annual dilated eye exam rates from 34% to 75%, but camera use peaked in the first 4 months and then steadily declined; moreover, only 27% of patients completed recommended in-person follow-up after abnormalities were identified, with transportation barriers identified as a major contributor [11,12]. In contrast, mobile pop-up eye clinics bring equipment and clinical personnel to primary care offices, providing in-person dilated eye examinations that can address several limitations of teleophthalmology, including the inability to perform dilated fundus examinations and the gap between screening and follow-up. However, far less evidence exists on the effectiveness of mobile vision vans for diabetic retinopathy screening in the United States [13-15].

Although prior studies have examined either teleophthalmology or mobile clinics in isolation, no study has examined the implementation of both modalities coexisting within the same rural primary care system. Using the Consolidated Framework for Implementation Research (CFIR), this mixed methods study examined how patients, clinic staff, primary care clinicians, and managers experienced teleophthalmology and pop-up eye clinics at 2 primary care clinics classified as “isolated rural” (Rural-Urban Commuting Area [RUCA] code 10) in upstate New York to identify multilevel barriers and facilitators and to understand how eye care can be more effectively integrated into and sustained within rural primary care settings.

Methods

Study Design and Settings

This mixed methods convergent implementation study was conducted at 2 university-affiliated rural primary care clinics (clinics A and B) serving predominantly low-income populations in the greater Rochester, New York region. The convergent design involved 2 data strands collected in parallel and analyzed independently: prospective semistructured interviews served as the primary data source for identifying implementation barriers and facilitators, while retrospective descriptive electronic health record (EHR) data characterized the patient population and eye care use patterns [16]. The 2 strands were integrated during interpretation. This study adhered to the Good Reporting of a Mixed Methods Study (GRAMMS) guidelines (Checklist 1) and the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Checklist 2) [17,18].

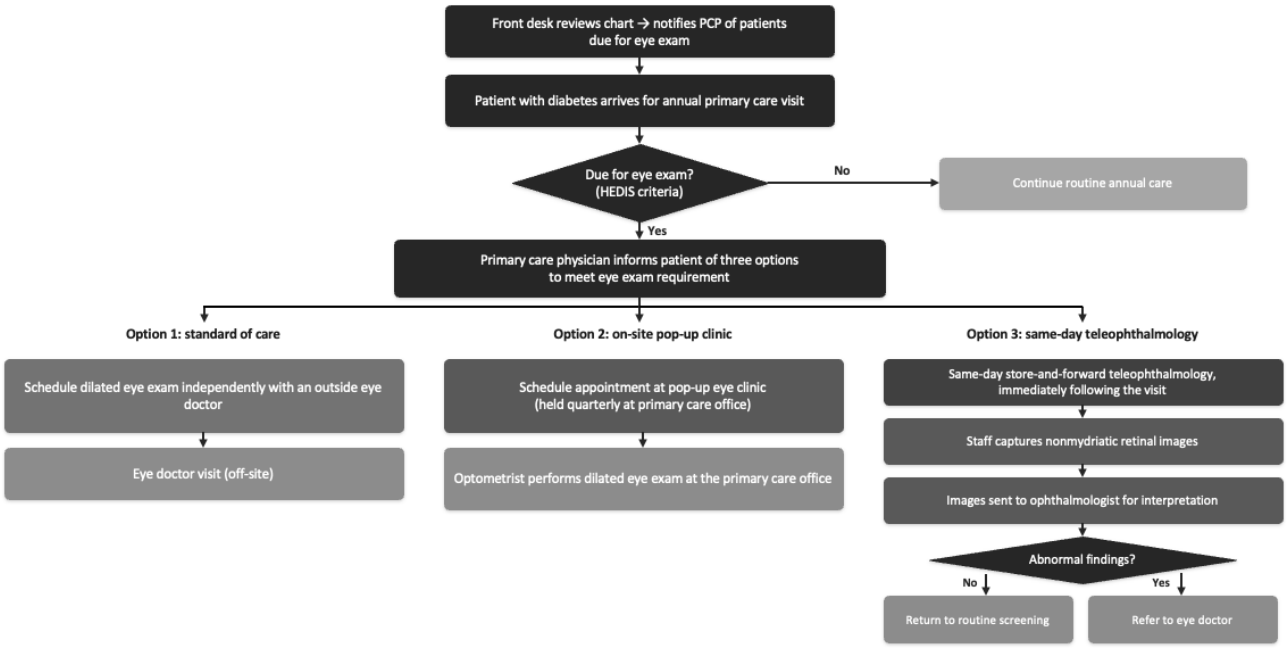
The clinics are classified as RUCA code 10 under the United States Department of Agriculture RUCA classification, indicating that the clinics are in isolated rural regions where the primary commuting flow is outside of any census-defined urbanized area [19]. Both clinics provide comprehensive primary care to adults with chronic conditions, including large numbers of patients with diabetes, and historically had lower baseline rates of annual diabetic eye examinations compared to system-wide targets. In collaboration with the university-based ophthalmology department, 2 on-site eye care delivery models were implemented: store-and-forward teleophthalmology embedded within primary care visits and quarterly pop-up eye clinics providing in-person dilated examinations. These interventions were introduced as strategies to increase access, reduce travel burden, and reach socially vulnerable patients who were not completing recommended annual eye examinations.

Clinic Workflow and Eye Care Programs

The current workflow for patients with diabetes due for an annual eye exam is represented in Figure 1. Front desk staff conduct previsit planning by reviewing incoming patients’ charts and notifying the primary care physician (PCP) of patients with diabetes who have not had an annual dilated eye exam in the previous year (Healthcare Effectiveness Data and Information Set [HEDIS] criteria for a diabetic eye exam [20]). At the primary care appointment, the PCP informs the patient of three options to meet their eye exam requirement: (1) independently schedule a dilated eye exam with an eye doctor, which represents the standard of care, (2) schedule an eye exam at the upcoming pop-up eye clinic held every 3 months at the primary care office, or (3) participate in store-and-forward teleophthalmology immediately following the appointment. Because options 2 and 3 are the on-site primary care–based interventions examined in this study, their workflows are described in detail below.

XSL•FO
RenderX

Figure 1. Current workflow. Options 2 and 3 are separate, unconnected pathways: no arrows link the pop-up eye clinic (option 2) and teleophthalmology (option 3) workflows, reflecting their current lack of integration. HEDIS: Healthcare Effectiveness Data and Information Set; PCP: primary care physician.



On the pop-up eye clinic day, a mobile vision van delivers eye exam equipment and personnel to an available room at the primary care office. A university-employed optometrist performs in-person comprehensive eye exams, including dilation to detect retinal and optic nerve diseases such as diabetic retinopathy, usually lasting around 45 minutes. Insurance billing for the pop-up eye clinic is done on the EHR platform in the same manner as at the university eye clinic. Ophthalmology and primary care staff both contact a list of patients with diabetes who need an eye exam to schedule them for an eye doctor appointment at the primary care clinic.

At a teleophthalmology visit, a trained primary care staff member assesses a patient's vision using a standard Snellen chart and uses a nonmydriatic NW400 Topcon digital retinal camera to take photographs of the patient's eyes. These images are sent to a university-based ophthalmologist for interpretation, who recommends the appropriate follow-up based on the findings from the teleophthalmology screening. These exams are billed to insurance using the 92,250 Current Procedural Terminology code [11].

Quantitative Data Collection

We used EHR data from 885 adult patients diagnosed with diabetes mellitus who received care at clinics A and B to describe the population eligible for eye care services. Patients were included if they had at least one annual visit with their PCP at one of the 2 clinics between January 1, 2022, and June 30, 2024; were aged 18 years or older; and had a diagnosis of diabetes. Variables extracted included demographic characteristics (age, race, and ethnicity), insurance type, social needs (transportation, housing instability, and food insecurity), health behaviors (tobacco use), comorbidities (hypertension, blood urea nitrogen, hemoglobin A1c, and low-density lipoprotein [LDL] levels), and eye pathology findings from the on-site examination groups.

Qualitative Data Collection

The CFIR guided both qualitative data collection and analysis [21]. For this study, four CFIR domains most relevant to the research questions were selected: (1) inner setting and individuals, (2) intervention characteristics, (3) implementation process, and (4) outer setting. Interview guides (Multimedia Appendix 1) and the coding framework were organized around these domains.

We conducted semistructured interviews with 20 patients with diabetes, sampling 5 from each of 4 subgroups: eye exam with an eye doctor, teleophthalmology, pop-up clinic, and no recent eye exam. Staff interviews were conducted with 8 clinic staff (2 PCPs and 2 staff members who had participated in teleophthalmology and pop-up eye clinics at each clinic) and 6 administrative staff (4 ophthalmology billing team members and 2 ophthalmology outreach program coordinators), for a total of 14 staff participants. No patient or staff member declined participation. All interviews were conducted by NGA, a medical student not involved in clinical care at the study sites, via telephone. Participants were informed that NGA was a medical student conducting research on eye care programs. Verbal consent was obtained, and all interviews were audio-recorded and professionally transcribed verbatim. Interviews lasted approximately 30 minutes. The interviewer recorded brief field notes after each interview to capture contextual observations and initial impressions; transcripts were not returned to participants for comment. Patients who were interviewed received US $25 for their participation. Interview guides were developed using these 4 CFIR domains to elicit perspectives on the implementation, integration, and sustainability of teleophthalmology and pop-up eye clinics (Multimedia Appendix 1).

XSL•FO
RenderX

Data Analysis

Quantitative data from 885 adult patients diagnosed with diabetes mellitus who received care at clinics A and B were used to describe the population eligible for eye care services. These descriptive data were used to characterize the study population and patterns of eye care use; no statistical comparisons for hypothesis testing were performed. Patients were categorized into three groups based on documented eye care use: (1) those who received a dilated eye exam by an eye doctor, (2) those who received an on-site eye exam at their primary care clinic through teleophthalmology or a pop-up eye clinic, and (3) those who had no eye exam in at least the last 2 years. We summarized demographic characteristics (age, race, and ethnicity), insurance type, social needs (transportation, housing instability, and food insecurity), health behaviors (tobacco use), and comorbidities (hypertension, blood urea nitrogen, hemoglobin A1c, and LDL levels) using means, medians, and proportions (Table 1). For patients in the pop-up eye clinic and teleophthalmology groups, we also calculated attendance rates and the frequencies of eye pathologies identified (cataract, diabetic retinopathy, and reduced visual acuity).

XSL•FO
RenderX

Table . Demographic data by group.

Characteristic	Total, N=885	Exam with eye doctor group, n=586	On-site exam at primary care group[a], n=91	No recent eye exam group[b], n=208
Primary care clinic, n (%)				
Clinic A	466 (52.7)	290 (49.5)	63 (69.2)	113 (54.3)
Clinic B	419 (47.3)	296 (50.5)	28 (30.8)	95 (45.7)
Age (y), mean (SD)	65.1 (13.4)	68 (13.2)	62.5 (12.9)	58.1 (11.2)
Sex, n (%)				
Male	474 (53.6)	314 (53.6)	51 (56.0)	109 (52.4)
Female	411 (46.4)	272 (46.4)	40 (43.9)	99 (47.6)
Race, n (%)				
White	813 (91.9)	541 (92.3)	79 (86.8)	193 (92.8)
Black	46 (5.2)	27 (4.6)	11 (12.1)	8 (3.8)
Other[c]	20 (2.3)	13 (2.2)	1 (1.1)	6 (2.9)
Unknown	6 (0.7)	5 (0.9)	0	1 (0.5)
Hispanic ethnicity, mean (SD)	25 (2.8)	20 (3.4)	1 (1.1)	4 (1.9)
Insurance, n (%)				
Commercial	329 (37.2)	183 (31.2)	33 (36.3)	113 (54.3)
Medicare Advantage	399 (45.1)	307 (52.4)	40 (43.9)	52 (25)
Medicare	96 (10.8)	66 (11.3)	10 (10.9)	20 (9.6)
Commercially managed Medicaid	53 (5.9)	25 (4.3)	8 (8.8)	20 (9.6)
Medicaid	2 (0.2)	1 (0.2)	0 (0)	1 (0.5)
Unknown	6 (0.7)	4 (0.7)	0 (0)	2 (0.9)
Transportation needs, n (%)				
No needs	494 (55.8)	326 (55.6)	42 (46.2)	126 (60.6)
Unmet needs	24 (2.7)	17 (2.9)	3 (3.3)	4 (1.9)
Unknown	367 (41.5)	243 (41.5)	46 (50.6)	78 (37.5)
Housing instability, n (%)				
Low risk	406 (45.9)	277 (47.3)	28 (30.8)	101 (48.6)
High risk	36 (4.1)	22 (3.8)	6 (6.6)	8 (3.8)
Unknown	443 (50.1)	287 (48.9)	57 (62.6)	99 (47.6)
Food insecurity, n (%)				
No insecurity	467 (52.8)	317 (54.1)	36 (39.6)	114 (54.8)
Insecurity present	46 (5.2)	24 (4.1)	6 (6.6)	16 (7.7)
Unknown	372 (42.0)	245 (41.8)	49 (53.8)	78 (37.5)
Tobacco use status, n (%)				
Current smoker	125 (14.1)	66 (11.3)	21 (23.1)	38 (18.3)
Past smoker	359 (40.6)	243 (41.5)	40 (43.9)	76 (36.5)
Never smoker	394 (44.5)	273 (46.6)	29 (31.9)	92 (44.2)
Unknown	7 (0.8)	4 (0.7)	1 (1.1)	2 (0.9)
Hypertension[d], n (%)	218 (24.6)	145 (24.7)	19 (20.9)	54 (25.9)
Median BUN[e] (range) (mg/dL)	17 (13-22)	18 (14-22)	16 (11-21)	16 (13-20)

XSL•FO
RenderX

Characteristic	Total, N=885	Exam with eye doctor group, n=586	On-site exam at primary care group[a], n=91	No recent eye exam group[b], n=208
Median HbA1c (range)	7.1 (6.4-7.9)	7.1 (6.4-7.8)	7.1 (6.4-7.9)	7.1 (6.3-8.1)
LDL[f] >100 mg/dL, n (%)	190 (21.5)	119 (20.3)	20 (21.9)	51 (24.5)

[a]Includes patients who received teleophthalmology or pop-up eye clinic exams.

[b]The patient did not meet Healthcare Effectiveness Data and Information Set (HEDIS) criteria for diabetic screening, defined as an eye exam every 2 years if there is no history of diabetic retinopathy, or an eye exam every year if there is a history of diabetic retinopathy.

[c]Includes American Indian, Alaska Native, and Asian.

[d]Systolic blood pressure >140 mm Hg or diastolic blood pressure >90 mm Hg, as defined by the American Diabetes Association.

[e]BUN: blood urea nitrogen.

[f]LDL: low-density lipoprotein.

Qualitative data were analyzed using reflexive thematic analysis as described by Braun and Clarke [22], informed by their 2023 recommendations for good practice in thematic analysis. The CFIR framework that guided interview development also served as a sensitizing framework during analysis, providing an initial organizing structure while allowing themes to be developed inductively from participants' accounts.

Three authors (NGA, RY-N, and RSR) independently read all transcripts for familiarization. NGA generated initial codes line-by-line across the full dataset, with RY-N and RSR independently coding a subset of transcripts. The 3 coders met to compare codes, discuss discrepancies, and iteratively refine a shared codebook maintained in Excel. Candidate themes were developed by collating codes into broader patterns of meaning and then reviewed against the coded data and full transcripts to ensure coherence and distinctiveness. Themes were defined, named, and organized into the 4 CFIR domains. Coding continued until no new themes were identified across successive transcripts, suggesting thematic sufficiency.

Reflexivity was addressed as follows: the research team included members with backgrounds in ophthalmology (RSR), implementation science and qualitative methods (RY-N), and medical education (NGA). NGA, who conducted all interviews, was a medical student not involved in clinical care at the study sites, which may have facilitated candid responses from both patients and staff. RSR, as the ophthalmologist who helped develop the teleophthalmology and pop-up clinic programs, brought deep contextual knowledge but also a potential investment in the programs' success; this was balanced by RY-N's methodological expertise and independent perspective during coding and theme development. The team regularly discussed how their respective positions might shape their interpretation and actively sought disconfirming evidence during analysis.

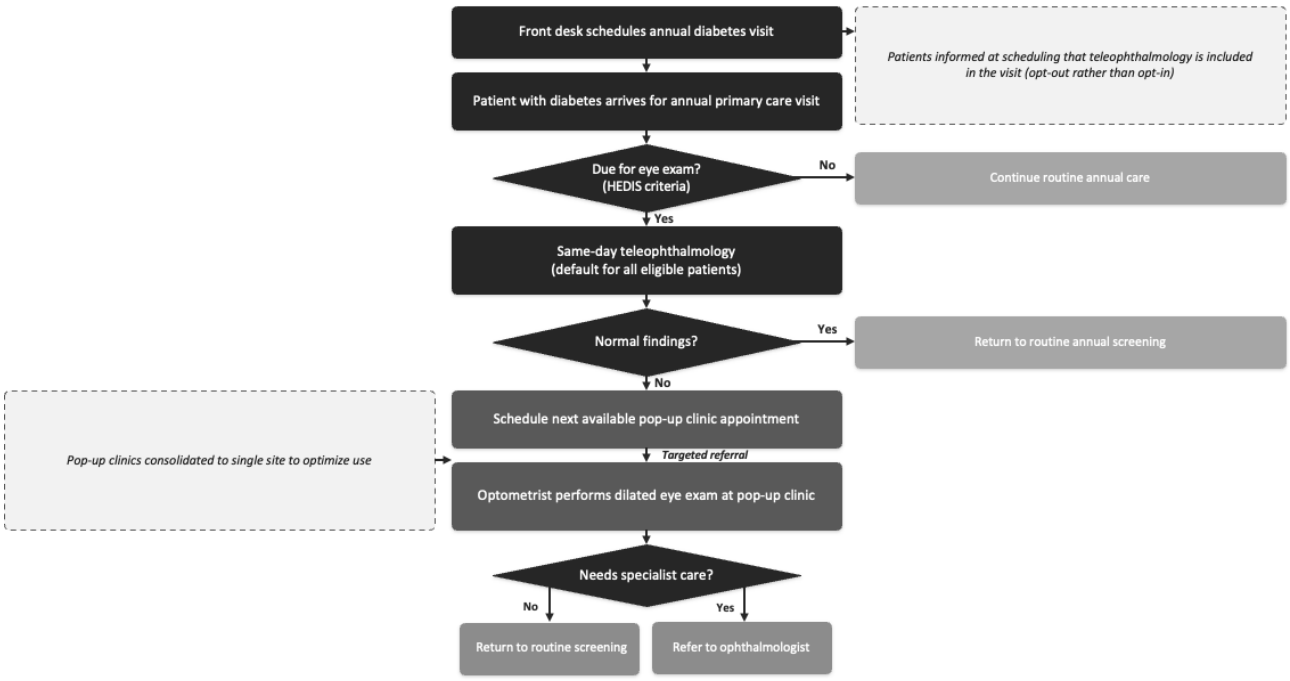
Several strategies were used to enhance the trustworthiness of the qualitative findings. Credibility was supported through investigator triangulation (3 coders with different disciplinary perspectives), purposive sampling across 4 patient subgroups and multiple staff roles, and the use of direct participant quotations to ground interpretations in the data. Dependability was maintained through a documented audit trail including the evolving codebook, meeting notes from consensus discussions, and decision logs for theme refinement. Confirmability was addressed through team-based analysis, reflexive discussion of researcher positionality, and systematic comparison of themes across participant groups. Transferability was supported by thick description of the rural clinical setting, patient population demographics, and implementation context, enabling readers to assess the applicability of findings to similar settings.

Current and Proposed Workflow Development

Two workflow diagrams were developed as analytic outputs of this study. First, the current eye care workflow (Figure 1) was constructed by mapping the existing clinical processes described by staff and administrative participants during semistructured interviews, triangulated with operational documents and direct observation of clinic procedures [23]. Second, the proposed integrated workflow (Figure 2) was developed through an iterative, stakeholder-driven synthesis process [24]. During qualitative analysis, specific implementation solutions and process redesign suggestions were identified as a distinct category of codes within the CFIR implementation process domain. These stakeholder-generated recommendations (spanning clinical staff [primary care physicians, registered nurses, and licensed practical nurses], administrative personnel [front desk staff and practice managers], visiting optometrists, and financial administrators) were extracted, mapped to the barriers and facilitators they were intended to address, and synthesized into a unified process model. The proposed workflow was refined through discussion among the research team (RSR, RY-N, and NGA) to ensure internal consistency, alignment with the qualitative findings, and feasibility within the constraints identified by participants. The final model was reviewed by 2 participating stakeholders (1 practice manager and 1 optometrist) for face validity.

XSL•FO
RenderX

Figure 2. Proposed integrated workflow. HEDIS: Healthcare Effectiveness Data and Information Set.



Ethical Considerations

This study was approved by the Institutional Review Board of the University of Rochester (RSRB00009510). Verbal informed consent was obtained from all interview participants prior to participation. All interviews were deidentified, and audio recordings were stored securely in accordance with institutional data security policies. Patient participants received US $25 in compensation for their time. The research described adhered to the tenets of the Declaration of Helsinki [25].

Results

Eye Care Use

Patient Demographics

Of the 885 patients with diabetes, approximately equal proportions were seen at the 2 clinics (Table 1). Overall, 586 (66.2%) received a dilated eye examination by an eye doctor and met HEDIS criteria (n=290 at clinic A and n=296 at clinic B). Of the remaining 299 patients who did not meet HEDIS criteria, 91 (30.4%) received an on-site primary care–based eye exam through teleophthalmology (n=39) or pop-up clinic (n=52), while 208 (69.6%) had not received an eye exam in the last 2 years.

Of 885 patients, most were older (mean age 65.1, SD 13.4 y) and self-identified as White (n=813, 91.9%), reflecting the demographic composition of the rural catchment area. All had health insurance, most commonly Medicare Advantage (n=399, 45.1%). Nearly half of the patients reported social needs: only 55.8% (n=494) reported no transportation needs, 45.9% (n=406) were at low risk for housing instability, and 52.8% (n=467) reported no food insecurity (Table 1).

Table 1 shows the descriptive characteristics of the study population by eye care use group. Patients in the on-site primary care eye exam group (n=91) were the most socially disadvantaged, with the lowest proportions reporting no transportation needs (n=42, 46.2%), low risk for housing instability (n=28, 30.8%), and no food insecurity (n=36, 39.6%). This group also had the highest prevalence of current smoking (n=21, 23.1%) and past smoking (n=40, 43.9%). Patients in the no recent eye exam group (n=208) were younger (mean age 58.1, SD 11.2 y) and had the highest rate of commercial insurance (n=113, 54.3%) but also the highest rates of hypertension (n=54, 25.9%) and elevated LDL levels (n=51, 24.5%) compared with the other groups.

Within the on-site primary care–based eye exam group, patients in the pop-up clinic and teleophthalmology subgroups were similar with respect to age (mean 63.3, SD 14.0 y vs mean 61.5, SD 11.4 y), sex (29/52, 55.8% vs 22/39, 56.4% male patients), race (43/52, 82.7% vs 36/39, 92.3% White), no transportation needs (24/52, 46.2% vs 18/39, 46.2%), low risk of housing instability (19/52, 36.5% vs 9/39, 23.1%), no food insecurity (21/52, 40.4% vs 15/39, 38.5%), and current smoking status (12/52, 23.1% vs 9/39, 23.1%).

Program Use and Reach

Over the 18-month study period, 14 half-day pop-up eye clinics were held at 2 rural primary care clinics. Each session offered 8 appointment slots (112 total), of which 89 (79.5%) were scheduled and 52 (46.4% of total; 58.4% of scheduled) were completed. Among the 39 patients in the teleophthalmology group, adequate image quality was obtained for 36 (92.3%). In 3 (7.7%) patients, poor image quality prevented grading for diabetic retinopathy, and these patients were referred for an in-person eye exam with an eye doctor within 3 months. Taken together, the pop-up eye clinic and teleophthalmology programs provided eye care for 30.4% (91/299) of patients who were due for an annual diabetic eye exam but had not seen an eye doctor.

XSL•FO
RenderX

Eye Pathology

Among the 52 patients seen in the pop-up eye clinics, where a comprehensive dilated exam was performed by an optometrist, a higher rate of eye pathology was identified, including cataracts diagnosed in 29 (55.8%), diabetic retinopathy in 11 (21.2%), and reduced visual acuity worse than 20/40 in 23 (44.2%). No eye pathology was identified in 8 (15.4%) pop-up clinic patients.

Among the 39 patients who received a nonmydriatic screening exam via teleophthalmology, cataract and diabetic retinopathy were each identified in 3 (7.7%) patients, macular pathology in 2 (5.1%), glaucoma in 1 (2.6%), and reduced visual acuity in 11 (28.2%). No eye pathology was identified in 24 (61.5%) teleophthalmology patients.

Facilitators of Implementation

Qualitative interviews with 20 patients and 14 staff members identified several facilitators of on-site eye care implementation (Table 2). Exact quotes can be found in Multimedia Appendix 2. Across all patient subgroups, proximity to the primary care office was the most frequently cited advantage, particularly for patients with transportation challenges: “It’s a long drive into the city, and I don’t drive anymore” [Teleophthalmology patient 5]. Closely related was the trust and familiarity that patients had developed with primary care staff over years of regular visits, which they contrasted with infrequent encounters at eye doctors’ offices: “They treat you like family, they don’t treat you like you’re just a patient” [Teleophthalmology patient 2].

XSL•FO
RenderX

Table . Summary of qualitative findings organized by Consolidated Framework for Implementation Research 2.0 domain.[a]

Theme	Barrier or facilitator	Brief finding
Individuals and inner setting		
Proximity of primary care office	Facilitator	Patients valued receiving eye care at a familiar, nearby location, reducing travel burden
Trust in primary care staff	Facilitator	Longstanding relationships with clinic staff increased willingness to participate
Assistance with scheduling referrals	Facilitator	Staff-assisted scheduling reduced patient burden of navigating follow-up appointments
Motivation and buy-in	Facilitator	Staff described the programs as meaningful and expressed pride in preventing vision loss
Limited staffing resources	Barrier	Short-staffing made it difficult to consistently assign trained personnel for teleophthalmology
Pop-up clinic underutilization	Barrier	Optometrists reported low patient volume at pop-up sessions relative to their regular clinics
Intervention characteristics		
“One-stop shop” integration	Facilitator	Patients preferred same-day, same-location eye care bundled with their diabetes visit
Patient perception of quality	Facilitator	Patients described on-site exams as thorough and comparable or superior to prior eye doctor visits
Continued need for glasses or specialty care	Barrier	Patients still required separate visits for eyeglasses and advanced eye disease management
Technical complexity of teleophthalmology	Barrier	Multistep camera and upload workflow created staff anxiety and inconsistent use
Implementation processes		
Workflow integration	Facilitator or barrier	Patients found same-day screening smooth and convenient, but lack of advance notification and competing transportation arrangements prevented some from participating
Advanced communication	Facilitator	Patients informed ahead of time were more willing to stay for same-day screening
Clinician-driven referral	Facilitator	Patients were more receptive when their primary care physician recommended on-site eye care
Implementation champions	Facilitator	Staff across roles proposed solutions and sustained program momentum despite challenges
Training gaps	Barrier	Only a subset of staff felt confident with the camera; errors caused delays
Coordination between staff	Barrier	Clinicians offered teleophthalmology inconsistently due to uncertainty about staff availability
Outer setting		
Insurance coverage	Mixed	Some patients had adequate coverage; others were uncertain about copays at on-site programs
Reimbursement disputes	Barrier	Billing disagreements with insurers over CPT[b] codes created financial uncertainty

[a]Representative quotes for each theme are provided in Multimedia Appendix 2.

[b]CPT: Current Procedural Terminology.

Patients across modalities described on-site eye care as a “one-stop shop” that integrated diabetes care and eye screening into a single visit, reducing the burden of scheduling and traveling to a separate appointment. Several patients noted that staff assistance with scheduling follow-up referrals was particularly valuable, as navigating the referral process independently was difficult. Patients who had used either modality expressed high satisfaction with the quality of care, with some describing on-site exams as more thorough than previous eye doctor visits.

Among staff, motivation and buy-in were strong facilitators. Nurses described the programs as meaningful ways to prevent

XSL•FO
RenderX

vision loss in patients who might otherwise "fall through the cracks," and front desk staff felt motivated to coordinate follow-up despite the extra workload. Implementation champions spanning clinical and administrative roles, including registered nurses, licensed practical nurses, front desk staff, practice managers, optometrists, and a financial administrator, played a critical role in sustaining the programs and proposing solutions to operational challenges. Clinician endorsement was a key facilitator: patients were more receptive when the recommendation came from their primary care clinician.

Barriers to Implementation

Several barriers limited the reach and efficiency of both programs (Table 2). For teleophthalmology, technical complexity was the most prominent barrier: staff described the camera workflow as involving many small steps, and missing a single step could prevent images from reaching the ophthalmologist. This was compounded by training gaps, with only a subset of staff at each clinic feeling confident operating the camera. One patient noted that their visit "took quite a while because neither of 'em knew how to do it" [Teleophthalmology patient 3]. Clinicians reported offering teleophthalmology inconsistently because of uncertainty about whether trained staff would be available, reflecting broader challenges with staff coordination.

For pop-up clinics, underutilization was the central barrier. Visiting optometrists reported that schedules were frequently unfilled, with 1 noting: "When I cancel to come to the pop-up clinic and there are only 2 patients on the schedule, it feels like a waste of time and money" [Optometrist 1]. Limited staffing resources affected both programs, as running eye care alongside routine primary care required pulling staff from other duties.

At the patient level, the continued need for eyeglasses and specialty care was identified as a limitation, as patients still had to travel elsewhere to purchase eyeglasses or receive treatment for complex eye disease. Regarding the outer setting, patients reported uncertainty about insurance coverage for on-site eye exams, and financial administrators described reimbursement disputes with insurers over Current Procedural Terminology code interpretation for teleophthalmology. Lack of advance communication about same-day teleophthalmology also led some patients to leave before screening because transportation arrangements had already been made.

Stakeholder-Proposed Integrated Workflow

Based on stakeholder recommendations identified through qualitative analysis, an integrated workflow was developed that links teleophthalmology and pop-up clinics into a single care pathway (Figure 2). In this model, all primary care staff are trained to perform teleophthalmology, which is positioned as a default component of annual diabetes visits rather than an optional add-on. At scheduling, front desk staff inform patients with diabetes that same-day teleophthalmology will follow their primary care visit. This default approach normalizes staff allocation and addresses clinician uncertainty about staff availability, identified as a barrier in the qualitative findings. Patients with abnormal teleophthalmology findings are scheduled directly into the next pop-up clinic for a comprehensive dilated exam, creating a predictable referral stream that addresses pop-up clinic underutilization. Patients with normal results return to routine screening intervals. Consistent with practice managers' recommendations, the model consolidates the 2 rural pop-up clinics into a single site to optimize optometrists' time and clinic space.

Discussion

Interpretation of Findings and Implications for Implementation

This mixed methods convergent study examined multilevel barriers and facilitators to implementing teleophthalmology and pop-up eye clinics for diabetic eye care in 2 isolated rural primary care clinics. On-site eye care programs preferentially reached socially vulnerable patients, that is, those with the highest rates of transportation needs, housing instability, food insecurity, and tobacco use, and detected substantial unmet eye pathology, including diabetic retinopathy in 21.2% (11/52) of pop-up clinic patients. However, operating the 2 modalities as separate, parallel programs created complementary inefficiencies: staff lacked confidence and consistent workflow support for teleophthalmology, while pop-up clinic appointment slots were underutilized. Stakeholder champions proposed an integrated workflow linking the 2 modalities into a single care pathway, which is presented in the *Results* section.

The finding that on-site eye care disproportionately reached socially disadvantaged patients suggests these programs may function as equity-targeting strategies rather than solely as screening interventions. Patients across all groups described the primary care office as a trusted, familiar "one-stop shop," reinforcing prior evidence that reducing travel burden and leveraging established primary care relationships facilitates screening engagement in rural communities [26,27]. The pop-up clinic diabetic retinopathy rate of 21.2% (11/52) is consistent with national estimates of 18% to 22% among US adults with diabetes [28,29]. The higher detection rates in pop-up clinics compared with teleophthalmology reflect the greater sensitivity of comprehensive dilated examinations for anterior segment disease and subtle retinal findings, rather than necessarily a difference in underlying disease burden [30,31].

The teleophthalmology diabetic retinopathy detection rate of 7.7% (3/39) was lower than the 11% reported in the prior University of Rochester teleophthalmology study using the same nonmydriatic approach, and lower than rates in other large-scale programs [11,32,33]. Beyond the small sample size, this may reflect inconsistent camera use by staff who lacked confidence with the technology. Leeman et al [11] reported that camera use peaked early and then declined steadily, and qualitative findings in the present study revealed similar patterns: staff described fear of "doing it wrong," while champions argued that routine use would make it "second nature." Infrequent use may reduce both screening volume and image quality, potentially lowering detection sensitivity.

A critical finding was the disconnect between patient and system perspectives: patients experienced teleophthalmology and pop-up clinics as a single way of "getting eye care at the primary

XSL•FO
RenderX

care office," while staff operationalized them as separate programs with distinct workflows. By examining these 2 coexisting modalities within the same rural primary care system, this study reveals how parallel implementation without integration creates complementary inefficiencies. The stakeholder-proposed integrated workflow addresses this tension by repositioning teleophthalmology as a default component of annual diabetes visits, shifting from an "opt-in" to an "opt-out" approach similar to the United Kingdom National Diabetic Eye Screening Programme [34]. This default design directly addresses 2 key barriers: clinician uncertainty about staff availability for teleophthalmology and failure to inform patients in advance about same-day screening. Linking abnormal findings to targeted pop-up clinic referrals creates a predictable referral stream that may improve pop-up use, which operated at only 58.4% (52/89) of scheduled capacity during the study period. Importantly, this model emerged from the convergent recommendations of frontline champions across clinical, administrative, and financial roles rather than from the research team, lending it practical credibility that may facilitate adoption. Economic evaluations support teleophthalmology's value in rural populations when adequately integrated and scaled [35], and the proposed workflow may provide the integration structure needed to realize this potential.

This study has several limitations. It was conducted in 2 clinics within a single university-affiliated health system in 1 rural region, which may limit generalizability to settings with different patient populations, staffing models, or reimbursement structures. The predominantly White population (813/855, 91.9%) reflects the catchment area demographics but limits applicability to more racially and ethnically diverse communities. The quantitative component was intentionally descriptive and not powered for hypothesis testing, and the number of patients in the teleophthalmology (n=39) and pop-up (n=52) groups was modest. Twenty patients and 14 staff members were interviewed, which constrains the range of perspectives captured. No patient experienced both modalities, so comparisons were based on individual experience and hypothetical descriptions rather than direct within-person comparisons. The proposed integrated workflow requires prospective evaluation before its effectiveness can be established.

Conclusions

Primary care–based eye programs in isolated rural settings can preferentially reach socially vulnerable patients with diabetes and detect substantial unmet eye pathology, positioning them as both screening and health equity interventions. However, implementing teleophthalmology and pop-up clinics as separate programs leads to predictable inefficiencies that threaten sustainability. The stakeholder-driven, integrated model that embeds teleophthalmology into routine diabetes care with targeted pop-up clinic referrals offers a pathway to align these complementary modalities. More broadly, this study illustrates that when multiple delivery innovations coexist within a single clinical setting, intentional integration, informed by the perspectives of frontline staff and patients, is essential to realize their combined potential. Future studies should evaluate whether this integrated workflow increases screening rates, staff proficiency, and follow-up completion and whether the model is transferable to other rural primary care systems.

Acknowledgments

No generative AI tools were used at any stage in the preparation of this manuscript.

Funding

This research was supported by the University of Rochester Office of Medical Education and by unrestricted funding to the University of Rochester Department of Ophthalmology from Research to Prevent Blindness. The teleophthalmology and pop-up clinic programs were supported by the Mother Cabrini Grant (FY19S311) and a Community Health Grant from Excellus BlueCross BlueShield. These funders had no role in the design or conduct of the study or in manuscript preparation.

Data Availability

The qualitative interview data generated during this study are not publicly available due to participant privacy and confidentiality protections under the institutional review board approval. Deidentified descriptive quantitative data may be made available upon reasonable request to the corresponding author.

Authors' Contributions

NGA drafted the original manuscript, conducted all patient and staff interviews, performed the qualitative coding and descriptive statistical analyses, and prepared the tables and figures.
VK assisted with the literature review, data organization, and manuscript preparation.
TCG coordinated patient recruitment, managed pop-up eye clinic logistics, and contributed to data collection and manuscript review.
SB coordinated community outreach for the teleophthalmology and pop-up eye clinic programs, facilitated patient navigation and scheduling, and contributed to manuscript review.
RY-N co-designed the study, developed the interview guides and qualitative coding framework, supervised the qualitative analysis, and critically revised the manuscript.

XSL•FO
RenderX

RSR conceived and designed the study, supervised all aspects of the research, codeveloped the interview guides and qualitative coding framework, and critically revised the manuscript for intellectual content.
All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 19 KB - diabetes_v11i1e106290_app1.docx]

Multimedia Appendix 2
Representative quotes organized by Consolidated Framework for Implementation Research 2.0 domain and theme.
[DOCX File, 21 KB - diabetes_v11i1e106290_app2.docx]

Checklist 1
GRAMMS checklist.
[DOCX File, 16 KB - diabetes_v11i1e106290_app3.docx]

Checklist 2
COREQ checklist.
[DOCX File, 21 KB - diabetes_v11i1e106290_app4.docx]

References

1. Shah J, Ahmed H, Weiss MB, et al. Eye care workforce density by provider type in urban and rural counties in the United States, 2010 to 2021. Graefes Arch Clin Exp Ophthalmol 2026 Jul 2. [doi: 10.1007/s00417-026-07356-4] [Medline: 42390569]
2. Soares RR, Mokhashi N, Sharpe J, et al. Patient accessibility to eye care in the United States. Ophthalmology 2023 Apr;130(4):354-360. [doi: 10.1016/j.ophtha.2022.11.017] [Medline: 36427602]
3. Feng PW, Ahluwalia A, Feng H, Adelman RA. National trends in the United States eye care workforce from 1995 to 2017. Am J Ophthalmol 2020 Oct;218:128-135. [doi: 10.1016/j.ajo.2020.05.018] [Medline: 32445703]
4. Berkowitz ST, Finn AP, Parikh R, Kuriyan AE, Patel S. Ophthalmology workforce projections in the United States, 2020 to 2035. Ophthalmology 2024 Feb;131(2):133-139. [doi: 10.1016/j.ophtha.2023.09.018] [Medline: 37739231]
5. Chin EK, Ventura BV, See KY, Seibles J, Park SS. Nonmydriatic fundus photography for teleophthalmology diabetic retinopathy screening in rural and urban clinics. Telemed J E Health 2014 Feb;20(2):102-108. [doi: 10.1089/tmj.2013.0042] [Medline: 24219153]
6. Simon LS, Davis ML, Medunjanin D, et al. National experience of technology-based eye care services: a comprehensive ophthalmology telemedicine initiative. Ophthalmology 2025 Apr;132(4):442-451. [doi: 10.1016/j.ophtha.2024.09.032] [Medline: 39368749]
7. Davis RM, Fowler S, Bellis K, Pockl J, Al Pakalnis V, Woldorf A. Telemedicine improves eye examination rates in individuals with diabetes: a model for eye-care delivery in underserved communities. Diabetes Care 2003 Aug;26(8):2476. [doi: 10.2337/diacare.26.8.2476] [Medline: 12882889]
8. Conlin PR, Fisch BM, Cavallerano AA, Cavallerano JD, Bursell SE, Aiello LM. Nonmydriatic teleretinal imaging improves adherence to annual eye examinations in patients with diabetes. J Rehabil Res Dev 2006;43(6):733-740. [doi: 10.1682/jrrd.2005.07.0117] [Medline: 17310422]
9. Li B, Powell AM, Hooper PL, Sheidow TG. Prospective evaluation of teleophthalmology in screening and recurrence monitoring of neovascular age-related macular degeneration: a randomized clinical trial. JAMA Ophthalmol 2015 Mar;133(3):276-282. [doi: 10.1001/jamaophthalmol.2014.5014] [Medline: 25473945]
10. Mansberger SL, Gleitsmann K, Gardiner S, et al. Comparing the effectiveness of telemedicine and traditional surveillance in providing diabetic retinopathy screening examinations: a randomized controlled trial. Telemed J E Health 2013 Dec;19(12):942-948. [doi: 10.1089/tmj.2012.0313] [Medline: 24102102]
11. Leeman S, Wang L, Johnson BA, Fortuna RJ, Ramchandran RS. Criteria-based assessment of a teleophthalmology diabetic retinopathy evaluation program in a primary care setting. Telemed J E Health 2022 Jun;28(6):865-872. [doi: 10.1089/tmj.2021.0064] [Medline: 34788158]
12. Ramchandran RS, Yilmaz S, Greaux E, Dozier A. Patient perceived value of teleophthalmology in an urban, low income US population with diabetes. PLoS One 2020;15(1):e0225300. [doi: 10.1371/journal.pone.0225300] [Medline: 31917793]
13. Lim A, Stewart J, Chui TY, et al. Prevalence and risk factors of diabetic retinopathy in a multi-racial underserved population. Ophthalmic Epidemiol 2008;15(6):402-409. [doi: 10.1080/09286580802435179] [Medline: 19065433]

XSL•FO
RenderX

14. Clark KK, Gutierrez J, Cody JR, Padilla BI. Implementation of diabetic retinopathy screening in adult patients with type 2 diabetes in a primary care setting. Clin Diabetes 2024;42(2):223-231. [doi: 10.2337/cd23-0032] [Medline: 38694241]
15. Nejatian MM, Sincari A, Alam K, Li I, Razavi H. Economic evaluations of eye care services for Indigenous populations in high-income countries: a scoping review. Int J Equity Health 2024 Nov 9;23(1):232. [doi: 10.1186/s12939-024-02307-z] [Medline: 39516782]
16. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE Publications; 2018.
17. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
18. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
19. Rural-Urban Commuting Area codes. USDA Economic Research Service, US Department of Agriculture. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes [accessed 2026-08-18]
20. HEDIS and performance measurement. National Committee for Quality Assurance (NCQA). URL: https://www.ncqa.org/hedis/ [accessed 2025-09-13]
21. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
22. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
23. Antonacci G, Lennox L, Barlow J, Evans L, Reed J. Process mapping in healthcare: a systematic review. BMC Health Serv Res 2021 Apr 14;21(1):342. [doi: 10.1186/s12913-021-06254-1] [Medline: 33853610]
24. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
25. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
26. Liu Y, Zupan NJ, Shiyanbola OO, et al. Factors influencing patient adherence with diabetic eye screening in rural communities: a qualitative study. PLoS One 2018;13(11):e0206742. [doi: 10.1371/journal.pone.0206742] [Medline: 30388172]
27. Snider MJE, Maa AY, Guyton AC, Park H, Hunt KJ, Pope C. Stakeholder perceptions affecting the implementation of teleophthalmology. BMC Health Serv Res 2022 Aug 25;22(1):1086. [doi: 10.1186/s12913-022-08386-4] [Medline: 36008809]
28. Lundeen EA, Burke-Conte Z, Rein DB, et al. Prevalence of diabetic retinopathy in the US in 2021. JAMA Ophthalmol 2023 Aug 1;141(8):747-754. [doi: 10.1001/jamaophthalmol.2023.2289] [Medline: 37318810]
29. Emmert R, Thompson M, Smith D, et al. Prevalence of diabetic retinopathy and dilated fundus examinations by metropolitan status from 2017-2021: an assessment of the Behavioral Risk Factor Surveillance System. Ophthalmic Epidemiol 2025 Oct;32(5):476-479. [doi: 10.1080/09286586.2024.2434247] [Medline: 39693586]
30. Weng CY, Maguire MG, Flaxel CJ, et al. Effectiveness of conventional digital fundus photography-based teleretinal screening for diabetic retinopathy and diabetic macular edema: a report by the American Academy of Ophthalmology. Ophthalmology 2024 Aug;131(8):927-942. [doi: 10.1016/j.ophtha.2024.02.017] [Medline: 38613533]
31. Vujosevic S, Aldington SJ, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol 2020 Apr;8(4):337-347. [doi: 10.1016/S2213-8587(19)30411-5] [Medline: 32113513]
32. Jani PD, Forbes L, Choudhury A, Preisser JS, Viera AJ, Garg S. Evaluation of diabetic retinal screening and factors for ophthalmology referral in a telemedicine network. JAMA Ophthalmol 2017 Jul 1;135(7):706-714. [doi: 10.1001/jamaophthalmol.2017.1150] [Medline: 28520833]
33. Bonilla-Escobar FJ, Ghobrial AI, Gallagher DS, Eller A, Waxman EL. Comprehensive insights into a decade-long journey: the evolution, impact, and human factors of an asynchronous telemedicine program for diabetic retinopathy screening in Pennsylvania, United States. PLoS One 2024;19(7):e0305586. [doi: 10.1371/journal.pone.0305586] [Medline: 38995899]
34. Diabetic eye screening: programme overview. GOV.UK. 2024. URL: https://www.gov.uk/guidance/diabetic-eye-screening-programme-overview [accessed 2025-01-12]
35. Sharafeldin N, Kawaguchi A, Sundaram A, et al. Review of economic evaluations of teleophthalmology as a screening strategy for chronic eye disease in adults. Br J Ophthalmol 2018 Nov;102(11):1485-1491. [doi: 10.1136/bjophthalmol-2017-311452] [Medline: 29680803]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record

XSL•FO
RenderX

GRAMMS: Good Reporting of a Mixed Methods Study
HEDIS: Healthcare Effectiveness Data and Information Set
LDL: low-density lipoprotein
PCP: primary care physician
RUCA: Rural-Urban Commuting Area

Edited by I Steenstra; submitted 05.Jul.2026; peer-reviewed by A Saxena, K Galaviz; revised version received 29.Jul.2026; accepted 11.Aug.2026; published 03.Sep.2026.

Please cite as:
Arya NG, Koltchine V, Gearinger TC, Buehler S, Yousefi-Nooraie R, Ramchandran RS
Implementing Diabetic Eye Care in Rural Primary Care Through Teleophthalmology and Pop-Up Clinics: Mixed Methods Study
JMIR Diabetes 2026;11:e106290
URL: https://diabetes.jmir.org/2026/1/e106290
doi:10.2196/106290

© Nisha G Arya, Victoria Koltchine, Talia C Gearinger, Sherianne Buehler, Reza Yousefi-Nooraie, Rajeev S Ramchandran. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 3.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Artificial Intelligence in Diabetic Kidney Disease Research: Bibliometric Analysis From 2006 to 2024

Xingyuan Li, MS; Liming Xiao, BS; Fenghao Yang, MS; Fang Liu, MD, PhD

Department of Nephrology, West China Hospital, Sichuan University, 37 Guoxue Alley, Wuhou District, Chengdu, Sichuan, China

Corresponding Author:
Fang Liu, MD, PhD
Department of Nephrology, West China Hospital, Sichuan University, 37 Guoxue Alley, Wuhou District, Chengdu, Sichuan, China

Abstract

Background: Diabetic kidney disease (DKD) is a major complication of diabetes and the leading cause of end-stage renal disease globally. Artificial intelligence (AI) technologies have shown increasing potential in DKD research for early detection, risk prediction, and disease management. However, the landscape of AI applications in this field remains incompletely mapped, especially in terms of collaboration networks, thematic evolution, and clinical translation.

Objective: This study aims to perform a comprehensive bibliometric and translational analysis of AI-related DKD research published between 2006 and 2024, identifying publication trends, research hotspots, key contributors, collaboration patterns, and the extent of clinical validation and explainability.

Methods: A systematic search of the Web of Science Core Collection was conducted to identify English-language original articles applying AI technologies to DKD. Articles were screened following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. Bibliometric visualization was performed using CiteSpace and VOSviewer to assess coauthorship, institutional and country collaboration, keyword evolution, and citation bursts. A qualitative review was conducted to evaluate clinical validation, model explainability, and real-world implementation.

Results: Out of 1158 retrieved records, 384 studies met the inclusion criteria. Global publications on AI in DKD increased rapidly after 2019. China led in publication volume, followed by the United States, India, and Iran. Keyword analysis showed a thematic transition from early biomarker and proteomic research to deep learning, clinical prediction models, and management tools. Despite methodological advances, few studies included external validation or explainability frameworks. Notable translational efforts included DeepMind's acute kidney injury predictor and a chronic kidney disease prediction model developed by Sumit, yet widespread real-world integration remains limited.

Conclusions: AI research in DKD has grown substantially over the past 2 decades, with expanding international collaboration and diversification of research themes. However, challenges persist in clinical applicability, model transparency, and global inclusivity. Future research should prioritize explainable AI, multicenter validation, and integration into clinical workflows to support effective translation of AI innovations into DKD care.

(JMIR Diabetes 2026;11:e72616) doi:10.2196/72616

KEYWORDS

artificial intelligence; diabetic kidney disease; bibliometric analysis; clinical validation; explainable AI; global collaboration

Introduction

Diabetic kidney disease (DKD) is the most prevalent microvascular complication of diabetes mellitus and a leading cause of end-stage renal disease globally, accounting for a substantial proportion of dialysis and transplantation burdens worldwide [1]. The pathophysiological progression of DKD is complex, often involving chronic hyperglycemia-induced glomerular injury, hemodynamic changes, inflammation, and fibrosis. Early-stage DKD is typically asymptomatic, and by the time clinical markers such as proteinuria or a decline in glomerular filtration rate become apparent, irreversible kidney damage may have already occurred [2]. Therefore, early detection and individualized risk stratification are essential for improving patient outcomes and alleviating long-term health care burdens.

In this context, artificial intelligence (AI) has emerged as a transformative approach in biomedical research and clinical practice. With capabilities in data-driven pattern recognition, predictive modeling, and real-time decision support, AI techniques—including machine learning, deep learning, and neural networks—have been increasingly explored to address key challenges in DKD research and management [3,4]. Applications range from biomarker discovery and disease classification to risk modeling and personalized treatment

XSL•FO
RenderX

optimization. Despite the growing enthusiasm for AI, there is wide variability in the methodological rigor, clinical applicability, and translational maturity of these studies.

While several narrative and systematic reviews have highlighted specific AI models used in nephrology, there remains a lack of comprehensive evaluation of how the field has evolved thematically over time, which countries and institutions are leading its development, how collaborative efforts are shaping knowledge production, and to what extent the proposed AI solutions are being validated and implemented in real-world clinical settings. Moreover, important dimensions such as model explainability, equity in global research representation, and translational readiness are often underexamined.

This study aims to address these gaps by conducting a bibliometric and translational landscape analysis of AI-related DKD research published from 2006 to 2024. By integrating quantitative bibliometric mapping with qualitative evaluation of translational attributes—including clinical validation, model transparency, and implementation potential—we aim to provide a comprehensive overview of this rapidly evolving field and offer insights to inform future research, clinical integration, and policy development.

Methods

Literature Search and Eligibility Criteria

A systematic literature search was conducted using the Web of Science Core Collection to identify studies related to the application of AI in DKD from January 1, 2006, to April 30, 2024. The search strategy included combinations of terms for DKD ("diabetic kidney disease," "diabetic nephropathy," "DKD," or "DN") and AI ("artificial intelligence," "machine learning," "deep learning," or "neural network"). Only English-language articles were considered. The search was limited to original research articles involving human-related data, excluding reviews, editorials, letters, conference abstracts, and purely experimental or theoretical reports without clinical relevance.

Eligible articles were those that applied AI techniques to DKD in a clinical, translational, or predictive context. Studies that involved image processing, signal detection, or statistical models unrelated to DKD-specific diagnostic or prognostic tasks were excluded. To ensure the reliability of inclusion, 2 reviewers (XL and FY) independently screened titles and abstracts for relevance, followed by full-text assessment. Discrepancies were resolved by consensus or consultation with a third reviewer (LX).

Bibliometric Mapping and Analysis Tools

Bibliometric data were exported from the Web of Science platform (Multimedia Appendix 1) and analyzed using CiteSpace (v6.1.R6) and VOSviewer (v1.6.18; Leiden University's Centre for Science and Technology Studies; Multimedia Appendix 2). These tools enabled visualization and quantification of publication trends, author and institutional productivity, international collaboration networks, and thematic keyword clusters. CiteSpace was used to generate timeline visualizations and detect emergent research topics through keyword burst detection. VOSviewer was applied to construct network maps illustrating coauthorship patterns and co-occurrence frequencies. Centrality scores and citation frequencies were used to identify influential authors, institutions, and countries within the research landscape.

Translational and Thematic Evaluation

In addition to bibliometric analysis, a qualitative assessment was performed to evaluate the translational significance of the included studies. This review focused on identifying whether AI models were externally validated or tested across different cohorts, whether explainable AI methods were incorporated, and whether any studies reported or discussed clinical integration or real-world implementation. Studies that mentioned the use of interpretability frameworks such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations) were noted. The presence of multicenter datasets, ethnically diverse populations, or cross-national data integration was also considered as indicators of generalizability and applicability. This dual approach—combining quantitative mapping with thematic content analysis—allowed for a multidimensional perspective on both the scientific growth and translational depth of AI research in DKD.

Ethical Considerations

This study involved no human participants, animals, or patient data, and therefore did not require ethical approval. The data used were retrieved from publicly available bibliographic databases and do not involve any sensitive or identifiable personal information.

Results

Study Selection

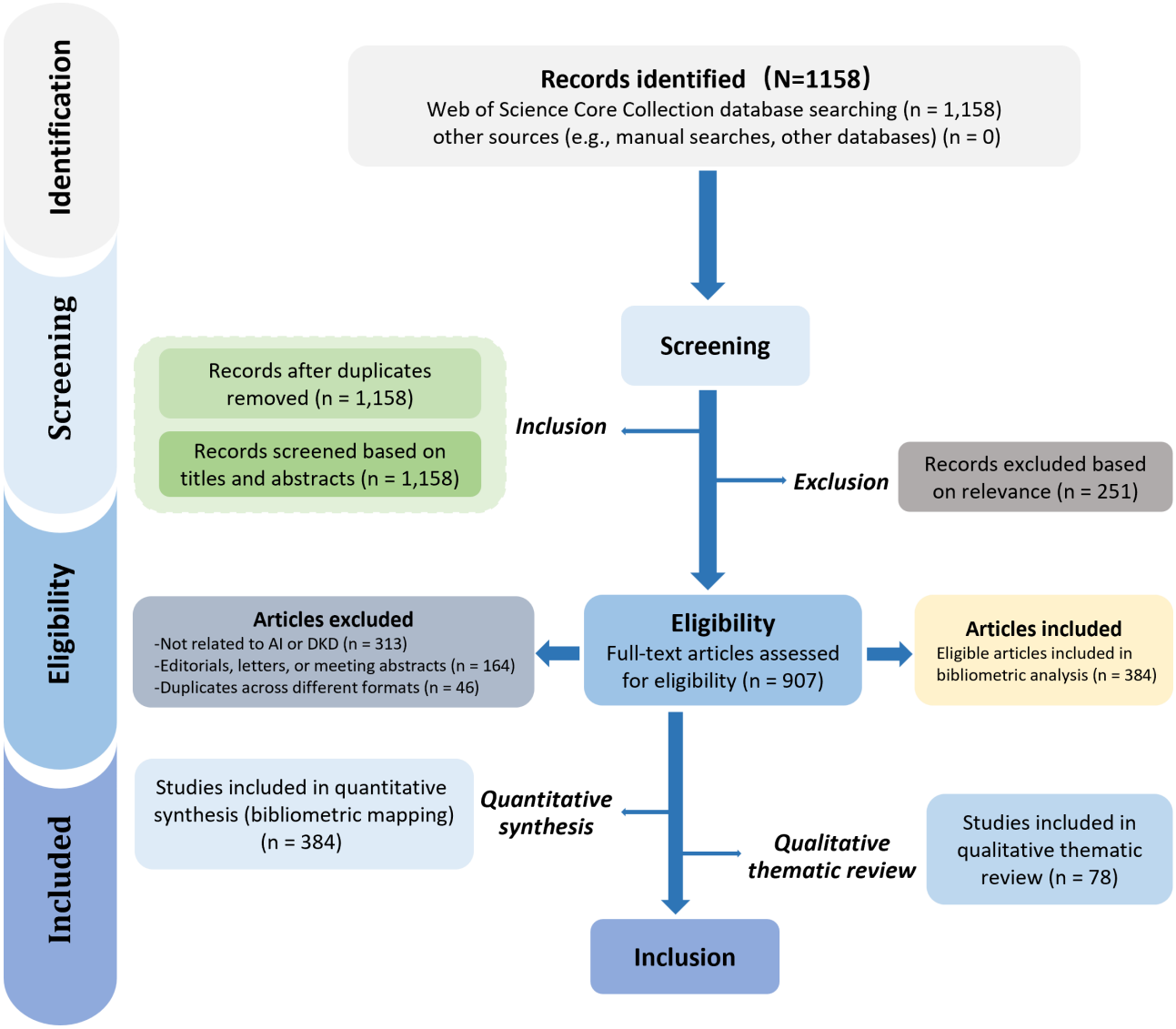
A total of 1158 records were initially identified following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 framework. After the removal of 0 duplicates, 1158 records were screened based on their titles and abstracts. Of these, 251 records were excluded as irrelevant. The remaining 907 full-text articles were assessed for eligibility, resulting in 384 articles included in the quantitative synthesis, and an additional 78 articles included in the qualitative thematic review. Ultimately, these articles were included in the subsequent bibliometric and qualitative synthesis.

Publication Growth Over Time

The global volume of publications related to AI in DKD remained low and relatively stagnant between 2006 and 2016. A notable increase in research output began in 2019, followed by a rapid rise during the years 2022 to 2024 (Figure 1). This pattern reflects the growing integration of AI techniques into biomedical research and the rising urgency of addressing DKD in the context of the global diabetes epidemic. The sharp upward trend in recent years suggests an increasing recognition of AI as a valuable tool for advancing DKD risk prediction, diagnosis, and management (Figure 2A).

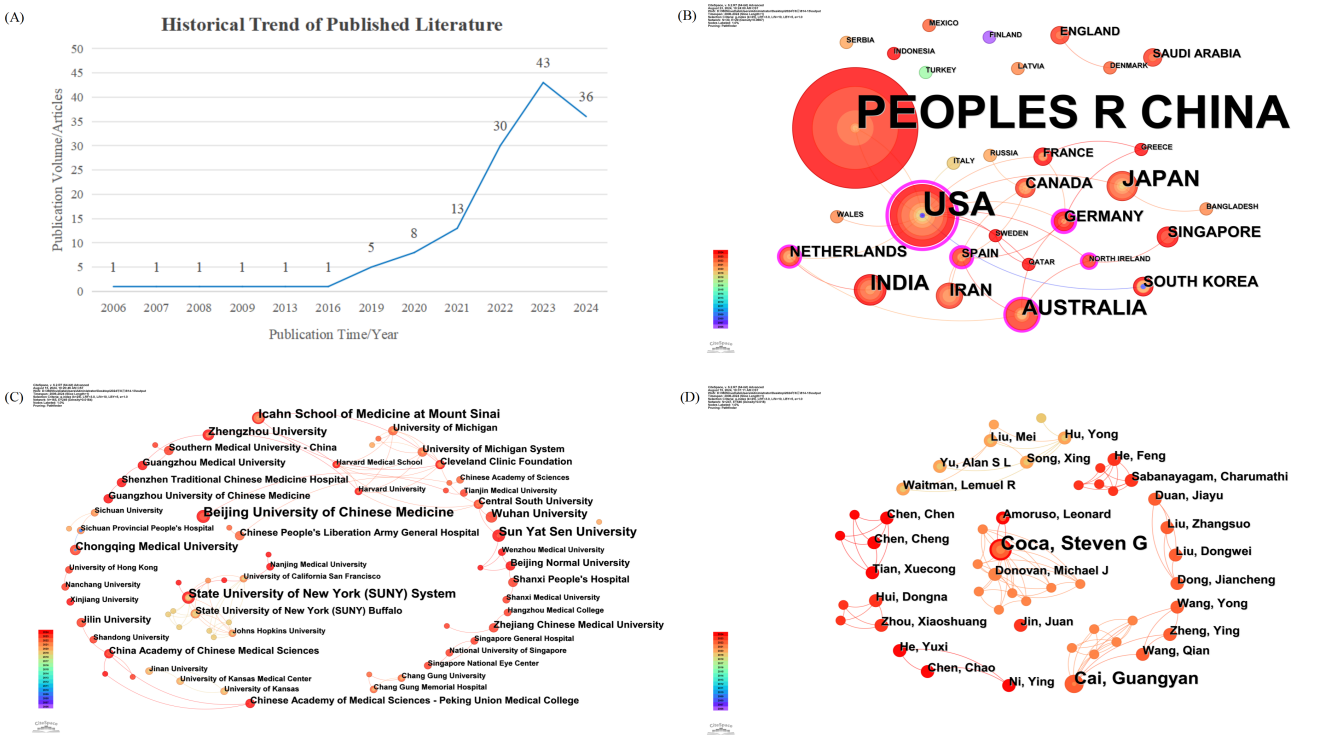
XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for literature screening. AI: artificial intelligence; DKD: diabetic kidney disease.



XSL•FO
RenderX

Figure 2. Analysis of the publication trends in artificial intelligence research on diabetic kidney disease from 2006 to 2024: (A) timeline of annual publications, (B) co-occurrence network of research countries, (C) co-occurrence network of research institutions, and (D) co-occurrence network of authors.



Geographic and Institutional Contributions

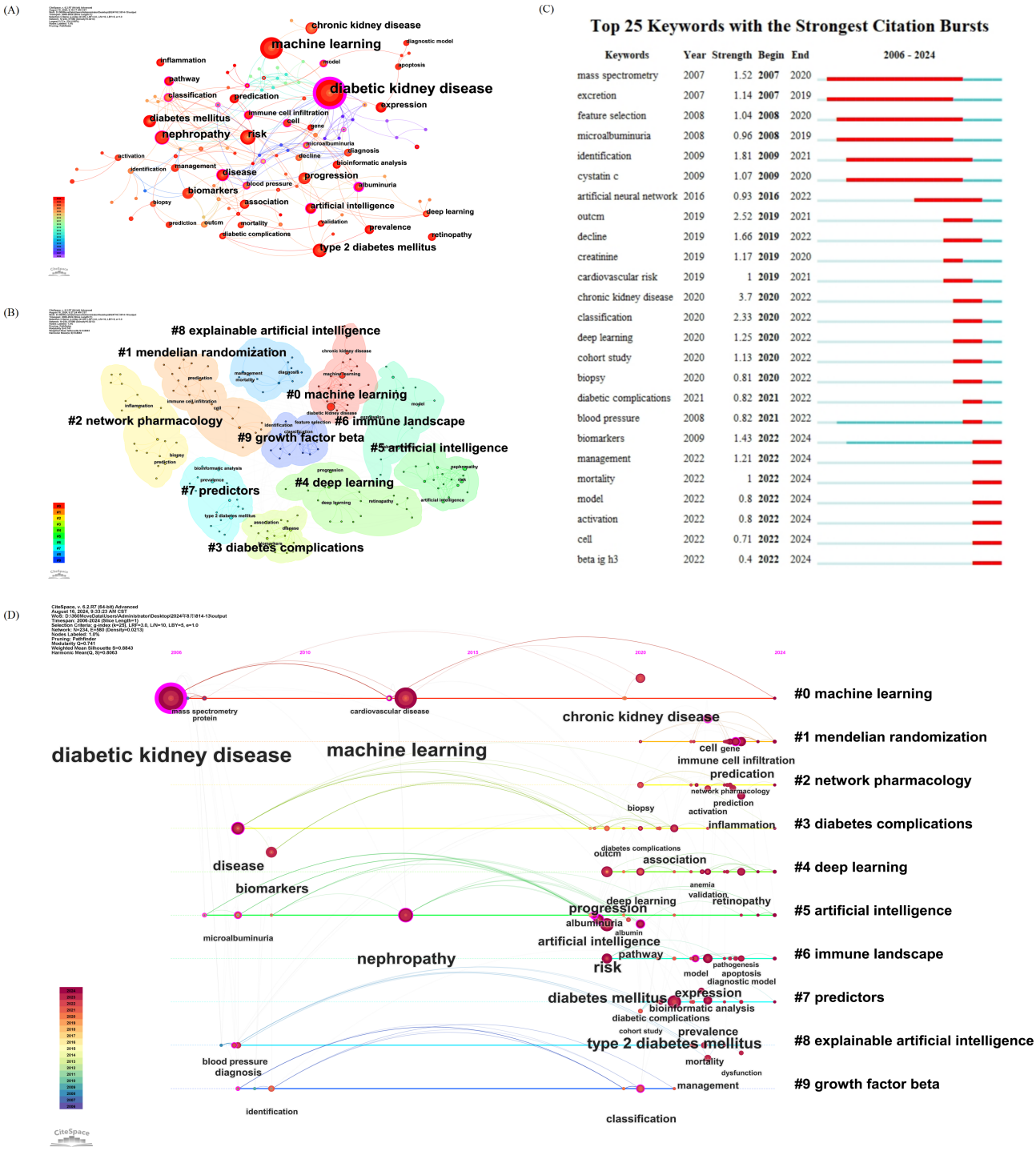
China emerged as the leading contributor in terms of publication volume, accounting for nearly half of all included studies. Key institutions such as Beijing University of Chinese Medicine, Sun Yat-sen University, and Central South University were among the most prolific. The United States ranked second, with prominent contributions from institutions such as the Icahn School of Medicine at Mount Sinai. India, Iran, and Australia also made notable contributions, reflecting a broader international interest in the intersection of AI and nephrology. Collaboration patterns showed that high-output countries often published independently, although intercontinental partnerships—particularly between East Asia, North America, and parts of Europe—have been increasing in frequency and visibility (Figure 2B-D).

Keyword Evolution and Research Hotspots

Analysis of keyword co-occurrence and burst terms revealed distinct phases in the thematic development of the field. During the early period (2006 - 2012), research was focused primarily on pathology, biomarker identification, and proteomic analysis, often using conventional statistical tools. Between 2013 and 2018, machine learning began to emerge as a prominent analytical method, with keywords such as “support vector machine” and “feature selection” gaining prominence. From 2019 onward, deep learning became a dominant theme, as reflected by the increasing frequency of terms such as “convolutional neural network,” “risk prediction,” and “decision support system.” Thematic clustering and citation bursts also indicated a growing interest in explainability, model integration, and individualized risk stratification, marking a shift toward clinical application and interpretability (Figure 3A-D).

XSL•FO
RenderX

Figure 3. Co-occurrence analysis of keywords in bibliometric studies: (A) keyword co-occurrence network, (B) keyword clustering, (C) keywords with the strongest citation bursts, and (D) timeline of keyword trends in artificial intelligence research on diabetic kidney disease from 2006 to 2024.



Collaboration Networks Among Authors and Institutions

Coauthorship network visualization demonstrated that the field remains highly fragmented, with a large number of small, loosely connected research groups. The most central nodes in the institutional network were located in China, the United States, and Singapore, reflecting both productivity and cross-institutional engagement. Although multicenter projects were occasionally identified, most AI models were developed and tested within single-center or regional datasets. Cross-national research, while increasing, often lacked shared validation protocols or harmonized data structures, limiting direct comparisons and large-scale model generalizability.

Model Validation, Explainability, and Translational Readiness

A review of the included studies showed that only a limited proportion of AI models underwent external validation using independent cohorts. Most models were based on retrospective data from a single institution or health system, with internal cross-validation as the primary method of evaluation. Very few studies implemented explainability frameworks such as SHAP or LIME, and even fewer offered insights into how model

XSL•FO
RenderX

outputs could be integrated into clinical decision-making processes. Notable exceptions included studies that incorporated prospective testing or demonstrated integration with electronic health records, although these remained rare. DeepMind's acute kidney injury prediction system, while not DKD-specific, was often cited as a prototype for nephrology-focused AI applications [5]. Similarly, Sumit's [6] deep learning–based model for chronic kidney disease risk prediction represented an example of real-world implementation relevant to diabetic populations. However, the lack of consistent attention to explainability, real-time integration, and regulatory considerations suggests that most AI-DKD research remains in a pretranslational stage.

Discussion

Principal Findings

This bibliometric and thematic analysis presents a comprehensive overview of research trends, international collaborations, and translational depth in the application of AI to DKD from 2006 to 2024. The temporal trend reveals a slow developmental phase lasting more than a decade, followed by a surge in research activity from 2019 onward. This acceleration corresponds with the broader adoption of AI in medicine and the urgent need for precision tools to combat the rising global burden of diabetes-related complications.

China and the United States have emerged as the primary contributors to this field, with China leading in publication quantity and institutional productivity. However, the dominance of single-country studies and weak international collaboration networks suggests a lack of unified global efforts in AI-DKD research. While some cross-border cooperation exists, it has not yet reached the level necessary to support large-scale model generalization or multiethnic validation. Future research should prioritize open data sharing, transnational model calibration, and harmonized validation protocols to promote reproducibility and clinical readiness across diverse populations.

Keyword analysis and thematic clustering indicate a clear evolution in research focus. Early studies emphasized molecular and pathological mechanisms of DKD, typically using traditional regression models or biomarker discovery tools. From 2015 onward, a shift occurred toward applying machine learning algorithms to structured clinical data, including risk prediction and feature selection. Since 2019, the field has seen a rapid proliferation of deep learning–based applications, especially convolutional neural networks for imaging and time-series data analysis. However, the transition from computational innovation to clinical implementation remains incomplete. Most studies prioritize model development and internal validation, while relatively few undertake real-world testing or prospective evaluation.

One major limitation identified is the scarcity of externally validated and clinically integrated AI models. Despite rapid algorithmic progress, few studies reach the level of clinical translation demonstrated by landmark systems such as DeepMind's acute kidney injury prediction algorithm, which was prospectively validated and tested in hospital settings [5]. Similarly, the work by Sumit [6], which developed and validated a deep learning model for predicting chronic kidney disease progression, represents an exemplar of real-world application. These examples underscore the importance of incorporating prospective design, external datasets, and health system integration early in the research pipeline to ensure that AI tools can transition beyond proof-of-concept stages.

Moreover, the "black box" nature of many AI models presents a significant barrier to clinical trust and regulatory approval. Although explainable artificial intelligence methods such as SHAP and LIME have been proposed and applied in other medical domains, they are seldom used in DKD-related research. This gap not only limits interpretability but also hinders integration into clinical workflows where explainability is essential for physician adoption and patient safety. The increasing interest in interpretable models and hybrid systems—combining clinical rules with machine learning outputs—may offer a promising path forward.

Another noteworthy observation is the underrepresentation of research from low- and middle-income countries, apart from China and India. Given the global prevalence of diabetes and its complications, this imbalance may reflect disparities in AI infrastructure, research funding, and access to large-scale clinical data. Efforts to democratize AI research—such as open-access datasets, international consortia, and capacity-building initiatives—are critical to avoid reinforcing health inequities through algorithmic bias.

Limitations and Future Work

This study also has limitations. The analysis was based solely on the Web of Science database, which, while comprehensive, may omit relevant studies indexed elsewhere, such as in Scopus or PubMed. The decision to focus on English-language articles may have further excluded important regional research. Additionally, bibliometric tools such as CiteSpace and VOSviewer, while effective in mapping research landscapes, cannot capture the full context or nuance of each study's methodological rigor or clinical relevance. Therefore, the qualitative thematic analysis presented here serves as a complementary lens, but further domain-specific review is warranted to assess clinical impact.

In conclusion, the field of AI in DKD is rapidly expanding, with increasing interest from diverse geographic regions and institutions. However, the translation of AI models into clinical nephrology practice remains limited. Future research should emphasize multicenter collaboration, external validation, and interpretability to close the gap between computational innovation and real-world impact. A systematic shift toward transparent, validated, and context-aware AI systems will be essential to unlock the full potential of AI in the management of DKD.

Conclusions

This study provides a comprehensive and multidimensional analysis of the research landscape at the intersection of AI and DKD. Through bibliometric visualization and thematic synthesis, we demonstrate that although the field has experienced substantial growth in recent years—particularly

XSL•FO
RenderX

with the application of deep learning technologies—the clinical translation of these innovations remains in its infancy. Most current research is confined to retrospective model development with limited external validation and minimal integration into real-world nephrology practice.

To advance the field, future efforts must prioritize methodological transparency, external validation using diverse populations, and the incorporation of explainable AI frameworks. Strengthening international collaboration and establishing multicenter consortia will be crucial for ensuring reproducibility and promoting equitable access to AI tools across health care settings. Additionally, regulatory and ethical considerations should be proactively addressed to support the safe deployment of AI in clinical decision-making.

In summary, while the promise of AI in DKD is evident, realizing its full potential will require a deliberate transition from algorithmic development to clinically meaningful, patient-centered applications. Bridging this translational gap is not only a technical challenge but also an opportunity to reshape chronic disease management in the era of intelligent medicine.

Acknowledgments

The authors thank all members of the biomedical informatics and nephrology research teams for their valuable discussions during the conceptualization of this study. We also acknowledge the developers of CiteSpace and VOSviewer for providing open-access tools that facilitated the bibliometric analysis. This study was supported by the Health Commission of Sichuan Province Program (21ZD001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Raw bibliometric data exported from the Web of Science Core Collection (CSV format, retrieved May 2024).
[PDF File, 19886 KB - diabetes_v11i1e72616_app1.pdf]

Multimedia Appendix 2
Analysis scripts and configuration settings used in CiteSpace (version 6.2.R6) and VOSviewer (version 1.6.19), provided in TXT and VOS formats.
[ZIP File, 22 KB - diabetes_v11i1e72616_app2.zip]

References

1. Umanath K, Lewis JB. Update on diabetic nephropathy: core curriculum 2018. Am J Kidney Dis 2018 Jun;71(6):884-895. [doi: 10.1053/j.ajkd.2017.10.026] [Medline: 29398179]
2. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. Clin J Am Soc Nephrol 2017 Dec 7;12(12):2032-2045. [doi: 10.2215/CJN.11491116] [Medline: 28522654]
3. Rodriguez-Romero V, Bergstrom RF, Decker BS, Lahu G, Vakilynejad M, Bies RR. Prediction of nephropathy in type 2 diabetes: an analysis of the ACCORD trial applying machine learning techniques. Clin Transl Sci 2019 Sep;12(5):519-528. [doi: 10.1111/cts.12647] [Medline: 31112000]
4. Makino M, Yoshimoto R, Ono M, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep 2019 Aug 14;9(1):11862. [doi: 10.1038/s41598-019-48263-5] [Medline: 31413285]
5. Ravizza S, Huschto T, Adamov A, et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. Nat Med 2019 Jan;25(1):57-59. [doi: 10.1038/s41591-018-0239-8] [Medline: 30617317]
6. Sumit K. Chronic kidney disease prediction using deep learning. Int J Sci Res Eng Manag 2025 May;9(5):1-4. [doi: 10.55041/IJSREM48077]

Abbreviations

AI: artificial intelligence
DKD: diabetic kidney disease
LIME: Local Interpretable Model-Agnostic Explanations
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SHAP: Shapley Additive Explanations

XSL•FO
RenderX

Edited by L Quinlan; submitted 13.Feb.2025; peer-reviewed by R Wen, S Banerjee, SR Polamuri; revised version received 22.Jul.2025; accepted 23.Jul.2025; published 09.Jan.2026.

Please cite as:
Li X, Xiao L, Yang F, Liu F
Artificial Intelligence in Diabetic Kidney Disease Research: Bibliometric Analysis From 2006 to 2024
JMIR Diabetes 2026;11:e72616
URL: https://diabetes.jmir.org/2026/1/e72616
doi:10.2196/72616

© Xingyuan Li, Liming Xiao, Fenghao Yang, Fang Liu. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 9.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing Quality, Readability, and Transparency of Online and AI-Generated Information on Type 2 Diabetes Mellitus: Cross-Sectional Exploratory Study

Sripradha Srinivasan[1], MBBCh; Ravanth Baskaran[1,2], MBBCh; Srinjay Mukhopadhyay[3,4], BSc, MBBCh; Mayank Patel[1], BM, DM

[1]University Hospital Southampton NHS Foundation Trust, Southampton, England, United Kingdom
[2]School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom
[3]National Heart and Lung Institute, Imperial College London, London, United Kingdom
[4]Imperial College Healthcare NHS Trust, The Bays, S Wharf Rd, London, England, United Kingdom

Corresponding Author:
Srinjay Mukhopadhyay, BSc, MBBCh
National Heart and Lung Institute, Imperial College London, London, United Kingdom

Abstract

Background: Type 2 diabetes mellitus (T2DM) affects approximately 590 million people worldwide, and its management relies heavily on patient education. With the emergence of online health information and artificial intelligence (AI) large language models, patients are increasingly sourcing medical information independently.

Objective: This study compared the quality, readability, and transparency of websites and AI-generated leaflets (AIGLs) related to T2DM.

Methods: Four predefined search terms (“type 2 diabetes,” “type 2 diabetes mellitus,” “T2DM,” and “adult diabetes”) were entered into 3 major search engines (Google, Yahoo, and Bing), and the top 20 search results were retrieved. AIGLs with patient information about T2DM were produced using a standardized prompt in 4 AI large language models (ChatGPT, Gemini, DeepSeek, and Grok). Information quality was assessed using the DISCERN score, calculated by 3 independent raters and ChatGPT. The *Journal of the American Medical Association (JAMA)* benchmarks were used to measure reliability and transparency. The Flesch-Kincaid Grade Level was used to determine readability.

Results: Seventy-five websites and 4 AIGLs were evaluated. Mean author-rated DISCERN scores were 42.6 (SD 11.3) for websites and 43.9 (SD 1.74) for AIGLs, corresponding to fair quality (DISCERN 41‑51). In contrast, ChatGPT-rated mean DISCERN scores were higher, with 58.5 (SD 11.5) for websites and 61.0 (SD 2.94) for AIGLs, corresponding to good quality (DISCERN 52‑63). Mean *JAMA* benchmark scores were 2.74 (SD 0.965) for websites, indicating moderate reliability (2-3 out of 4 points), whereas all AIGLs scored 0 out of 4 points. Mean Flesch-Kincaid Grade Level scores for websites were 8.67 (SD 2.23) and 8.30 (SD 1.92) for AIGLs, corresponding to an eighth- to ninth-grade comprehension level. Spearman rank correlation demonstrated minimal variability among the 3 independent raters but showed a significant difference between ChatGPT and the 3 independent raters.

Conclusions: Given the high prevalence of T2DM, both websites and AIGLs demonstrated suboptimal quality, readability, and transparency. Increasing patient reliance on digital health information calls for improved readability standards and stronger safeguards for AI-generated content. Both websites and AIGLs require an eighth- to ninth-grade comprehension level, far above the average reading age of 9 years in the United Kingdom (fourth- to fifth-grade level). This reduces the accessibility of online health information. The landscape of medical consultations is evolving, with patients increasingly presenting with preconceived notions based on online health information; hence, health care professionals should adapt to this shift.

(JMIR Diabetes 2026;11:e93822) doi:10.2196/93822

KEYWORDS

artificial intelligence; AI; ChatGPT; digital health; online patient information; type 2 diabetes mellitus

Introduction

Type 2 diabetes mellitus (T2DM) is a metabolic disorder characterized by chronic hyperglycemia resulting from insulin resistance [1]. By 2025, the International Diabetes Federation estimated that 590 million people worldwide would have diabetes, with approximately 90% of cases attributed to T2DM [2]. Initial management of T2DM typically involves dietary

XSL•FO
RenderX

modification and weight loss [3], followed by pharmacological treatment to reduce blood glucose through various mechanisms [4]. Inadequate management increases the risk of complications affecting the cardiovascular, renal, peripheral vascular, ophthalmic, hepatic, or neurological systems [5]. Effective management depends on patient understanding and adherence to therapy, making patient education a fundamental component of T2DM care [6].

The proliferation of digital technologies has led to a marked increase in the number of patients seeking health-related information online [7]. Population-based studies indicate that online health information–seeking behavior rose from approximately 61% in 2008 to 74% in 2017 [7]. In diabetes care, a cross-sectional study in the United States found that 64.5% of 1744 individuals with diabetes engaged in online health information seeking [8]. Digital platforms for accessing health information include websites, search engines, social media, live streaming platforms, and, more recently, artificial intelligence (AI) large language models (LLMs; [9]). This increased reliance on online sources is primarily attributed to their convenience, accessibility, and immediacy [10].

AI LLMs have gained popularity for health information access due to their personalized, empathetic, and conversational interfaces [11]. Research on generative AI for diabetes self-management has demonstrated high patient engagement, attributed to reduced information overload and the delivery of unbiased, transparent, individualized, and evidence-based responses [12]. However, the integration of AI LLMs into health information delivery raises concerns regarding ethical issues, regulatory risks associated with hyperpersonalization, and privacy implications [11,12]. Furthermore, patient safety may be compromised by medical hallucinations, where AI LLMs disseminate misinformation due to inadequate calibration and knowledge gaps [13].

While the quality of traditional online health resources and AI LLM–generated information has been evaluated in previous studies, direct comparisons between these 2 resource types remain limited [14,15]. As patient reliance on AI-generated health content increases, systematic evaluation of its quality relative to established online resources is essential. This study aimed to compare traditional online resources with AI-generated information on T2DM by (1) assessing the overall quality of the resources, (2) evaluating the readability of both information sources, and (3) comparing their reliability and transparency.

Methods

Overview

This study used a cross-sectional observational design to evaluate the quality, readability, and transparency of online and AI-generated information related to T2DM. Four predefined search terms were entered into 3 commonly used search engines, and the top 20 search results were retrieved. The 4 search terms were “type 2 diabetes,” “type 2 diabetes mellitus,” “T2DM,” and “adult diabetes.” The search engines used were Google, Yahoo, and Bing. Searches were conducted in Google Chrome using an incognito window, with history and cache cleared prior to the searches. The search was conducted on December 12, 2025. Sponsored results and advertisements were excluded when collating the top 20 search results. The search results were subsequently screened for duplicates and eligibility. In addition to duplicate entries, resources were excluded due to access restrictions (n=7), descriptions of health services rather than educational content (n=2), information about type 1 diabetes mellitus (T1DM; n=1), video-only content (n=2), and textbook purchase links (n=2). Websites presenting information on both T1DM and T2DM were eligible for inclusion, but only material relevant to T2DM was extracted and analyzed. When AI evaluated websites containing information on both T1DM and T2DM, the prompt explicitly instructed it to assess only the T2DM-specific content.

The websites that met our inclusion criteria were classified as public, research-based, hospital-related, or educational (Table 1). Interrater agreement was achieved on the predefined criteria for each category.

Table . Definitions of website categories.

Category	Definition
Public websites	Consumer-facing health information written for the public using nontechnical language (eg, NHS and Mayo Clinic)
Research-based websites	Sources that publish scientific or academic content, such as peer-reviewed articles, research summaries, or technical reports intended for academic or clinical audiences
Hospital-related websites	Websites produced by hospitals, health care trusts, or clinical institutions, typically containing clinical guidelines, care protocols, or service-specific information
Educational websites	Instructional or textbook-style resources designed for learners (eg, high-school, undergraduate, or early-training health care students) rather than the public

AI-generated leaflets (AIGLs) were created and collated using 4 different AI LLMs: ChatGPT (OpenAI), Gemini (Google LLC), DeepSeek (DeepSeek AI), and Grok (xAI; Figure 1). The same prompt was used across all 4 AI LLMs to create the AIGL: “Create a patient leaflet outlining information regarding type 2 diabetes.” The AI outputs were generated from a single run per AI LLM. Each assessment was conducted in a new chat with all prior history cleared, and a new account was created for each LLM to ensure no previous interactions influenced the results.

XSL•FO
RenderX

The quality of information was assessed using the DISCERN instrument, a valid, evidence-based tool to evaluate the quality of written consumer health information regarding treatment choices [16,17]. The DISCERN instrument consists of 16 items, each scored on a 5-point Likert scale, yielding a maximum score of 80, with higher scores indicating higher quality (Table 2). Each website and AIGL was independently assessed using the DISCERN criteria by 3 independent raters. Additionally, DISCERN scores were generated using ChatGPT to evaluate the same resources, enabling comparison between human and AI LLM assessments. ChatGPT was used to represent AI-rated DISCERN scores as it is the most widely used conversational LLM in health care research, with 92% of all health-related LLM studies using ChatGPT [18]. The prompt provided to ChatGPT stated, "Rate this website [insert website URL] using the DISCERN score. Give it a total score out of 80." ChatGPT evaluated the full webpage by analyzing the content accessible through the entered URL. When ChatGPT-rated websites containing T1DM and T2DM information, the prompt used was "Rate this website [insert website URL] using the DISCERN score. Give it a total score out of 80. Assess only the information related to type 2 diabetes."

Figure 1. Flowchart displaying the large language models used to generate health information leaflets.

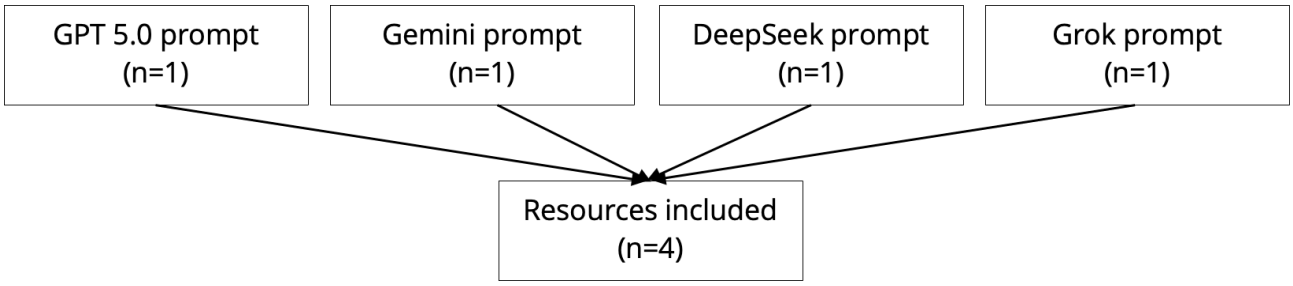


Table . DISCERN scores and corresponding quality levels.

DISCERN score	Quality
64 - 80	Excellent
52 - 63	Good
41 - 51	Fair
30 - 40	Poor
16 - 29	Very poor

Readability was judged using the Flesch-Kincaid Grade Level (FKGL). This metric is widely used in health literacy research to evaluate the accessibility of patient education materials [19]. This measure estimates the educational level required to comprehend a text based on sentence length and word complexity [20]. The score was derived using an online calculator in which the text was entered, and the formula automatically generated the result. The FKGL corresponds to a US school grade level, with lower grade levels indicating greater readability (Table 3).

Table . Domain categories and corresponding Flesch-Kincaid Grade Level (FKGL), US grade level, and UK educational equivalent.

Descriptive categories	FKGL (US educational stages)	FKGL (UK educational stages)
Very easy	Fifth grade	Year 6
Easy	Sixth grade	Year 7
Fairly easy	Seventh grade	Year 8
Plain English or standard	Eighth and ninth grades	Years 9 and 10
Fairly difficult	10th to 12th grades	Years 11 to 13
Difficult	College	Undergraduate degree student
Very difficult	College graduate	Undergraduate degree graduate

Lastly, the *Journal of the American Medical Association (JAMA)* benchmark criteria were used to explore the transparency and reliability of the resources. The *JAMA* score considers 4 indicators of information reliability: authorship, attribution, disclosure, and currency [21]. Each criterion was recorded as present or absent with a maximum score of 4. The higher the score, the greater its reliability (Table 4).

Normality of continuous variables was assessed using the Shapiro-Wilk test. Both author-rated and AI-rated DISCERN, *JAMA* benchmarks, and FKGL scores demonstrated a P value<.05, indicating a significant deviation from normality. The Mann-Whitney U test was used to compare scores between websites and AIGLs. The Wilcoxon signed-rank test was conducted to compare author-rated and AI-rated DISCERN scores. Spearman rank correlation coefficient was used to

XSL•FO
RenderX

compare variability between author-rated DISCERN scores. Spearman rank correlation was also used to compare ChatGPT-rated and author-rated DISCERN scores for websites. Scatterplots and best-fit lines of DISCERN, *JAMA* benchmarks, and FKGL scores were created based on the top 10 websites' positions in search engine results. This identified trends between a website's position and its quality and readability, with slope and *P* value depicting the best-fit line's significance from nonzero values displayed. A scatterplot with an associated best-fit line was used to graphically represent and identify any association between quality (DISCERN score) and readability (FKGL), with its slope and *P* value reported. Data collection was conducted in Microsoft Excel (16.109.2), and data analysis was conducted on GraphPad Prism (11.0.2; GraphPad Software LLC).

Table . *Journal of the American Medical Association (JAMA)* scores and corresponding quality levels.

JAMA benchmark score	Reliability
0 - 1	Poor reliability
2	Moderate reliability
3 - 4	High reliability

Ethical Considerations

An internal research ethics evaluation determined that formal institutional review board, research ethics board, or research ethics committee approval was not required, as the study involved no human participants, patient data, or identifiable information. As no human participants were included, informed consent was not applicable. Privacy and confidentiality were inherently maintained, and no participant compensation was involved.

Results

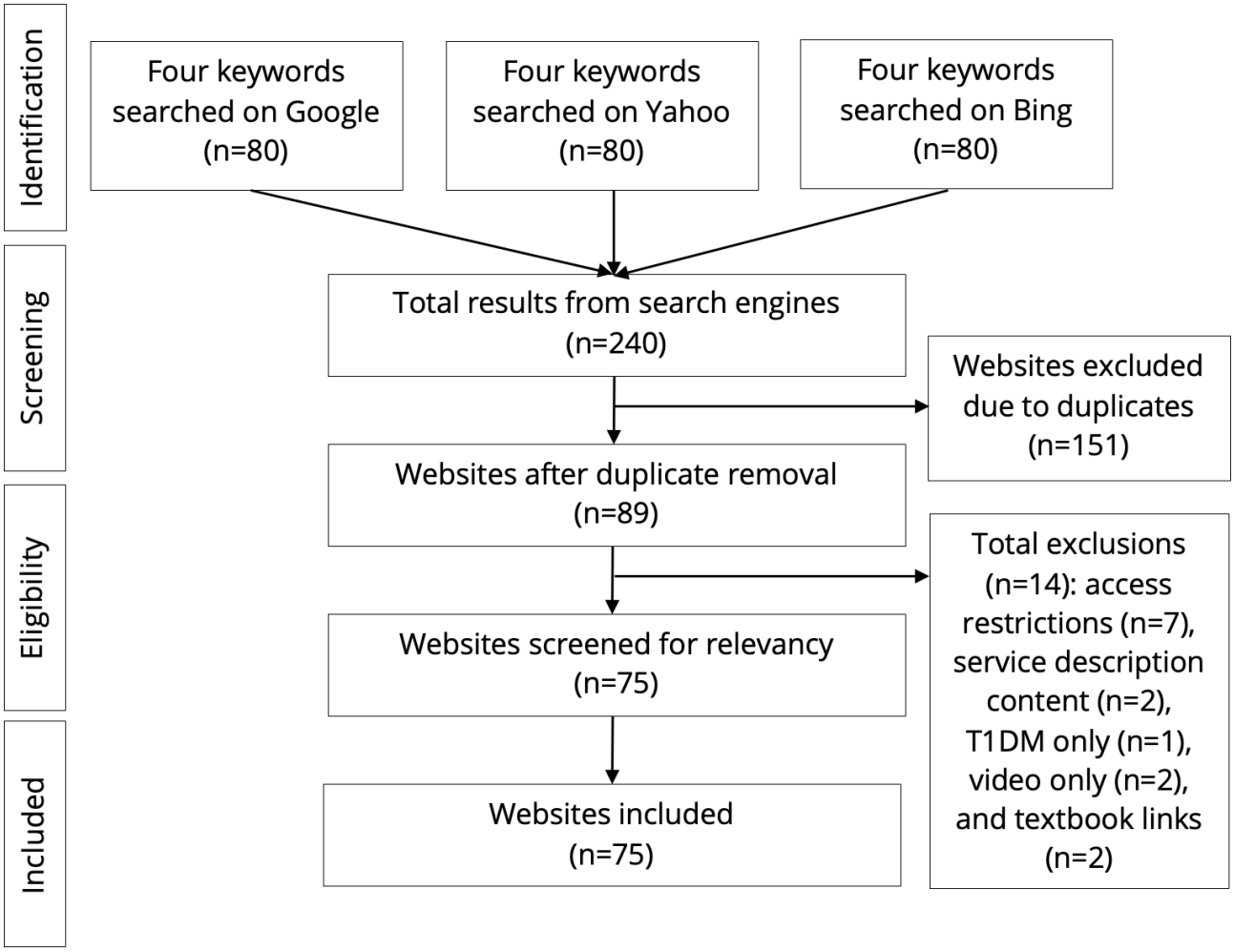
Demographics

The study selection process is shown in Figure 2. A total of 75 websites were included in the analysis. Of these, 50 (66.7%) originated in the United Kingdom, 19 (26.3%) in the United States, 5 (6.7%) were of multicountry origin, and 1 (1.3%) was from Spain. Regarding their type, 46 (61.3%) were public information websites, 18 (24%) were research-based, 9 (12%) were hospital-related, 1 (1.3%) was educational, and 1 (1.3%) was from a textbook.

Four AIGLs were produced using ChatGPT, Gemini, DeepSeek, and Grok, and analyzed alongside the websites.

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flowchart illustrating identification, screening, and exclusion of websites. T1DM: type 1 diabetes mellitus.



Quality of Information—DISCERN Scores for Websites vs AIGLs

There was no significant difference in the overall author-rated DISCERN scores between traditional websites and AIGLs (Table 5). Cohen *d* was 0.12 for this, suggesting a minimal effect.

Table . Comparison of author-rated DISCERN scores between websites and artificial intelligence–generated leaflets (AIGLs).

Source	DISCERN score, mean (SD)	*P* value
Websites (n=75)	42.6 (11.3)	.51
AIGLs (n=4)	43.9 (1.74)	.51

Quality of Information—Human vs AI DISCERN Ratings

A significant difference ($P<.001$) was found between the author-rated and AI-rated DISCERN scores for websites (n=75), with AI consistently assigning higher quality scores than a human assessor (Table 6).

In contrast, no significant difference was observed between author and AI rankings of AIGLs (n=4) themselves (Table 6).

XSL•FO
RenderX

Table . Comparison of author-rated and artificial intelligence (AI)–rated DISCERN scores between websites and AI-generated leaflets (AIGLs).

Source	DISCERN score, mean (SD)	*P* value
Websites (n=75)		<.001
Author-rated	42.6 (11.3)	
AI-rated	58.5 (11.5)	
AIGLs (n=4)		.13
Author-rated	43.9 (1.74)	
AI-rated	61.0 (2.94)	

Quality of Information—AI-Rated DISCERN Comparisons

When assessed exclusively by AI, no significant difference (*P*=.81) was found between the DISCERN scores of websites and AIGL (Table 7).

Table . Comparison of artificial intelligence (AI)–rated DISCERN scores between websites and AI-generated leaflets (AIGLs).

Source	DISCERN score, mean (SD)	*P* value
Websites (n=75)	58.5 (11.5)	.81
AIGLs (n=4)	61.0 (2.94)	.81

Reliability and Transparency—*JAMA* Benchmarks

A difference in reliability was identified between the average *JAMA* benchmark scores for websites and AIGLs. Websites achieved substantially higher *JAMA* benchmark scores, while AIGLs scored zero across all criteria (Table 8). The effect size was not calculated as AIGLs scored 0 for the means and standard deviations.

Table . Comparison of *Journal of the American Medical Association (JAMA)* benchmark scores between websites and artificial intelligence–generated leaflets (AIGLs)[a].

Source	*JAMA* benchmark score, mean (SD)
Websites (n=75)	2.74 (0.965)
AIGLs (n=4)	0 (0)

[a]No associated *P* value, as all AIGLs scored 0 on the *JAMA* benchmark.

Readability—FKGL

No significant difference (*P*=.77) was seen between the FKGL scores of the websites and the AIGLs (Table 9). Both websites and AIGLs required the educational level of an eighth to ninth grader to understand the texts. Cohen *d* was calculated to be 0.13, suggesting a minimal effect.

Table . Comparison of Flesch-Kincaid Grade Level (FKGL) scores between websites and artificial intelligence–generated leaflets (AIGLs).

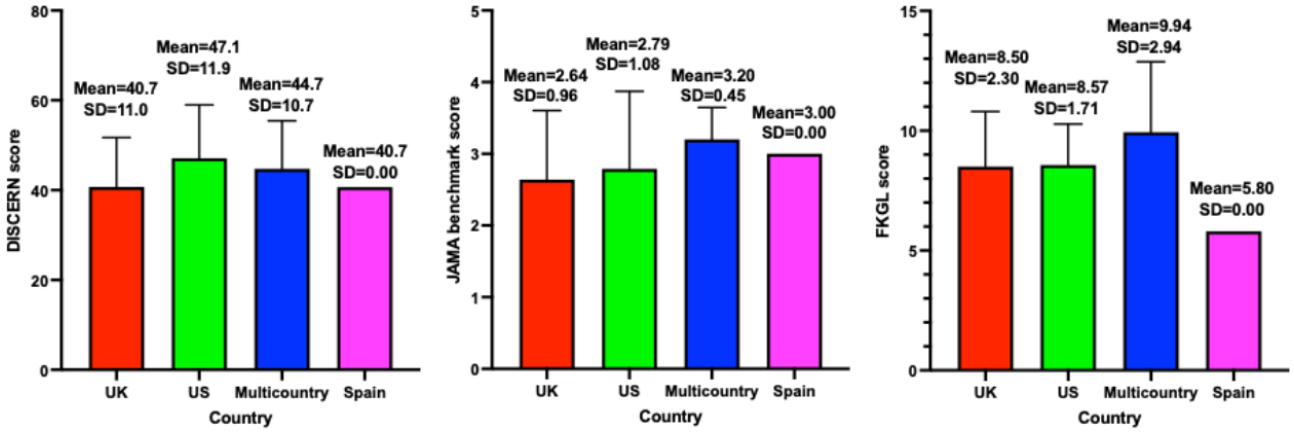
Source	FKGl score, mean (SD)	*P* value
Websites (n=75)	8.67 (2.23)	.77
AIGLs (n=4)	8.30 (1.92)	.77

Associations—Country of Origin

Websites originating from the United States achieved higher quality scores, as indicated by higher DISCERN scores. The highest reliability and transparency, as reflected by the *JAMA* benchmark score, was attained by the “multicountry” category of websites. The most readable websites were also observed in the “multicountry” category (Figure 3).

XSL•FO
RenderX

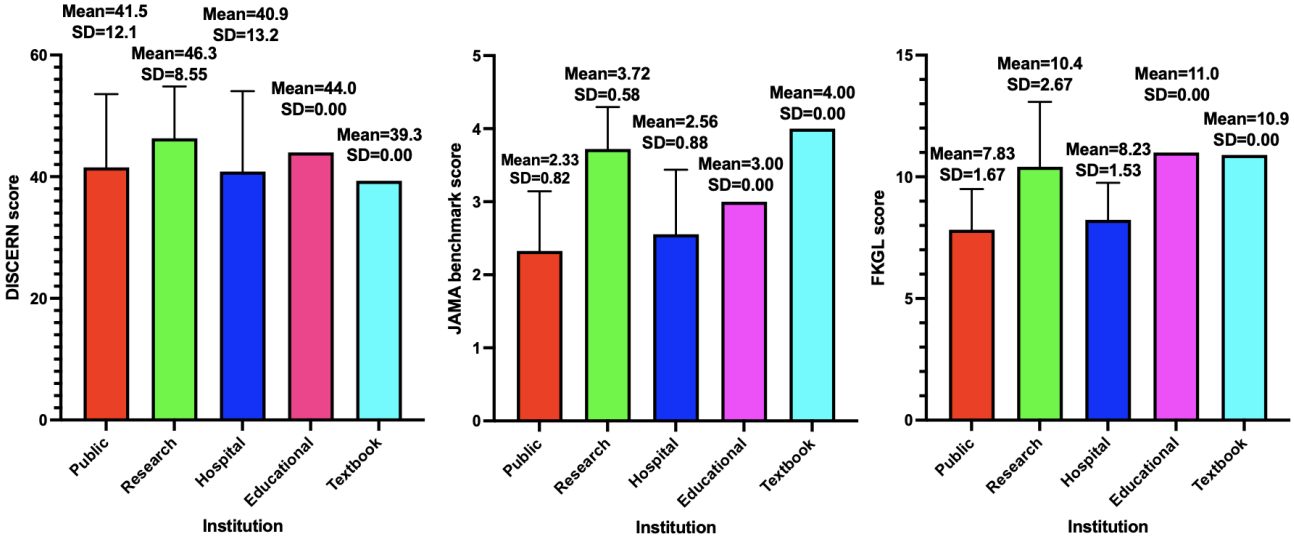
Figure 3. Graphical representation of the means and SDs of DISCERN, *Journal of the American Medical Association (JAMA)* benchmark, and Flesch-Kincaid Grade Level (FKGL) scores based on the origin countries of the websites.



Associations—Institutions

Research-based websites (n=18) had the highest DISCERN scores, while the textbook (n=1) had the highest *JAMA* benchmark scores, followed by research articles (n=18). The educational resource website (n=1) and the textbook website (n=1) had the highest FKGL scores. Variability in quality and readability was observed across institutions (Figure 4).

Figure 4. Graphical representation of the means and SDs of DISCERN, *Journal of the American Medical Association (JAMA)* benchmark and Flesch-Kincaid Grade Level (FKGL) scores based on the institutional origin of the website.

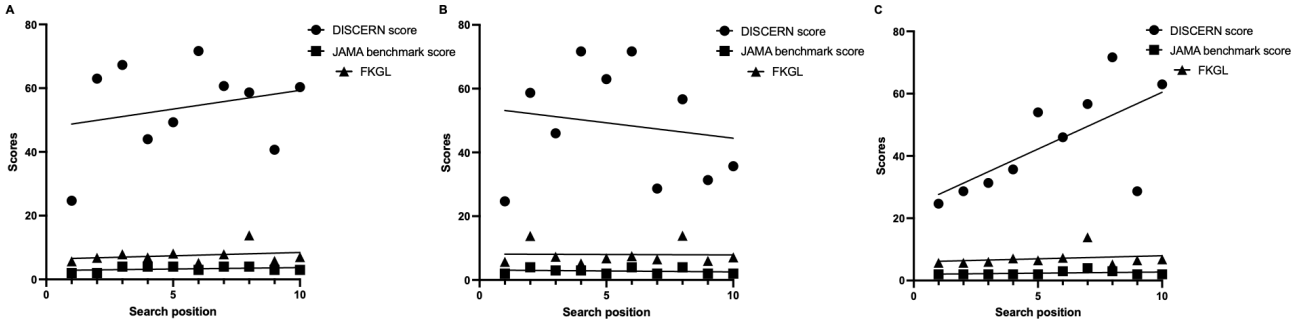


Search Engine Ranking

DISCERN, *JAMA* benchmark, and FKGL scores demonstrated a nonsignificant positive trend within the Google search engine. In contrast, all 3 metrics showed a nonsignificant negative trend in the Yahoo search engine. For Bing, the DISCERN score exhibited a significant positive trend, indicating that lower-quality websites tended to appear higher in the search results. The *JAMA* benchmark and FKGL scores, however, showed nonsignificant positive trends in Bing (Figure 5).

XSL•FO
RenderX

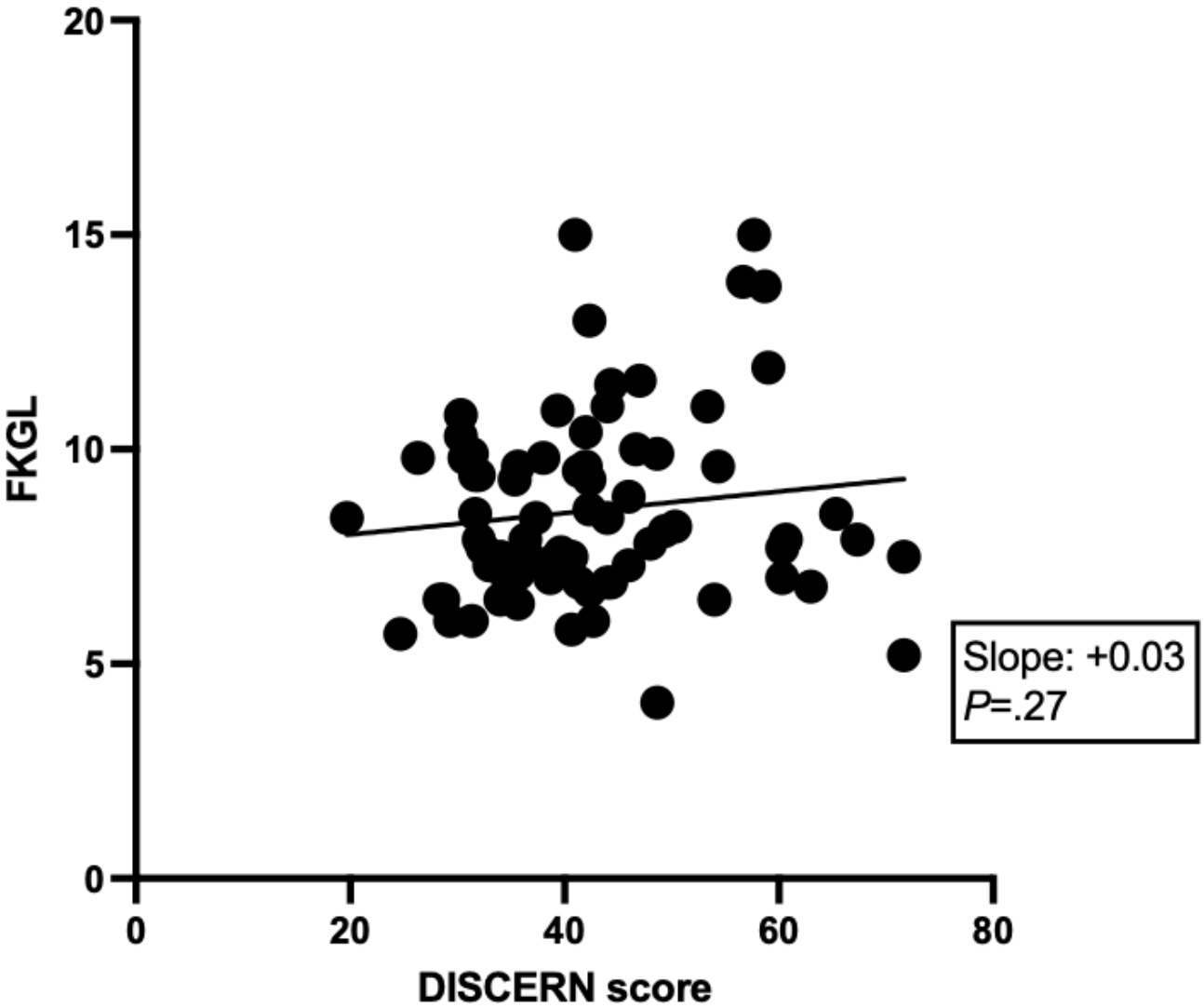
Figure 5. Scatterplots and best-fit lines of DISCERN, *Journal of the American Medical Association (JAMA)* benchmark, and Flesch-Kincaid Grade Level (FKGL) scores for Google, Yahoo, and Bing, based on the websites' positions in the top 10 search engine results. (A) scatterplot of correlation of DISCERN (slope: +1.175; *P*=.53), *JAMA* benchmark (slope: +0.091; *P* >.99), and FKGL (slope: +0.208; *P*=.59) scores and search position on Google, (B) scatterplot of correlation of DISCERN (slope: −0.996; *P*=.65), *JAMA* benchmark (slope: −0.061; *P*=.58), and FKGL scores (slope: −0.027; *P*=.94) and search position on Yahoo, and (C) scatterplot of correlation of DISCERN (slope: +3.65; *P*=.04), *JAMA* benchmark (slope: +0.073; *P*=.38), and FKGL scores (slope: +0.198; *P*=.50) and search position on Bing. Associated slopes and *P* value representing significance from nonzero slope are represented in tables.



Scoring Correlations

Comparisons between website quality (DISCERN score) and readability (FKGL) were performed to identify correlations. No significant correlation was found between website quality and readability, suggesting that higher-quality websites were neither more nor less difficult to read (Figure 6).

Figure 6. Scatterplot with best-fit line comparing DISCERN scores and Flesch-Kincaid Grade Level (FKGL) with associated slope and *P* value representing significance from nonzero slope.



XSL•FO
RenderX

Interrater Agreement

There was a strong and significant correlation among the DISCERN score raters, SS, RB, and SM (Table 10). This suggested minimal variability in the DISCERN scoring process between the raters. The sample size for AIGLs (n=4) was too small to generate a valid Spearman rank correlation.

When comparing the author-rated DISCERN scores for websites with those rated by ChatGPT, a notable discrepancy was observed between the 2 raters, indicating that AI applies DISCERN scoring differently than human raters (Table 11).

Table . Variability assessed using Spearman rank correlation analysis of DISCERN scores between authors SS, RB, and SM.

Source and rater comparison	Spearman *r*	*P* value
Websites		
SS vs RB	0.90	<.0001
SS vs SM	0.88	<.0001
RB vs SM	0.92	<.0001

Table . Variability calculated by conducting a Spearman rank correlation test between ChatGPT and author-rated DISCERN scores.

Source and rater comparison	Spearman *r*	*P* value
Websites		
ChatGPT vs SS	0.218	.06
ChatGPT vs RB	0.208	.07
ChatGPT vs SM	0.169	.15
Author average vs ChatGPT	0.204	.08

Discussion

Principal Findings

Although both online health information and AI have been extensively studied, this is, to our knowledge, the first study to compare them in the context of T2DM.

The overall quality of information provided by websites and AIGLs did not differ significantly when assessed by human evaluators. However, both sources received average author-rated DISCERN scores categorized as “fair quality.” In contrast, ChatGPT assigned both AIGLs and traditional websites an average score categorized as “good quality.” Additionally, the comparison of interrater agreement between authors showed no significant variability according to the Spearman rank correlation, whereas the comparison of DISCERN scores between authors and AI revealed significant categorical variability. This discrepancy raises the possibility that AI may overestimate resource quality by obscuring qualitative differences that human reviewers can identify. Supporting this, a study on the decoy effect found that AI LLMs were more likely to assign higher ratings to misinformation, particularly when false information was more attention-grabbing, whereas humans were better at identifying such inaccuracies [22]. Human assessments, however, also exhibited a decoy effect due to prior knowledge biases [22]. Therefore, while AI LLMs can generate unbiased information, they remain less effective than humans at eliminating errors.

AI-generated content that is factually incorrect yet confidently presented is referred to as a “hallucination” [23]. The phenomenon of hallucinations in newer AI LLMs has been widely discussed [24]. Although recent versions of ChatGPT provide more accurate information, the hallucination rate remains high, at 48%, representing a 15% increase from the previous model [25]. Advancements in AI LLMs have reduced overt errors, making it increasingly difficult for users to distinguish between factual information and confabulations [26]. This challenge is compounded by the assertiveness with which AI LLMs present misinformation. The 6 leading AI LLMs (GPT 4o, Llama-3-70B, Claude 3.5 Sonnet, Gemini, Claude 3 Opus, and GPT 4) demonstrated a mean confidence score of 69.4% when incorrect and 72.5% when correct [27]. Furthermore, AI LLMs have been shown to align with user opinions, even when these are factually inaccurate [28]. The combination of assertiveness and inadequate error elimination poses significant risks to patient safety, as errors may be presented as facts [29].

The rate of AI hallucinations may be reduced if AI LLMs consistently cite their sources. However, all AIGLs produced by the 4 different AI LLMs in this study received a mean *JAMA* benchmark score of 0, indicating a lack of authorship, disclosure, and safeguarding. This finding should be interpreted with caution, as the *JAMA* benchmark was developed to assess traditional web-based health information and may not fully capture the characteristics of AI-generated resources. Consequently, the low scores likely reflect a structural mismatch between the evaluation tool and AI-generated content rather than indicating that the information provided is inherently unreliable. Nevertheless, the absence of these transparency features remains an important limitation because it reduces users’ ability to verify information. In contrast, traditional websites demonstrated greater reliability and transparency, with a mean *JAMA* benchmark score of 2.7. These results are consistent with existing literature, which finds that AI LLMs

XSL•FO
RenderX

routinely lack transparency compared with traditional online resources [30]. Additionally, the opaque back end processing, screening, and exclusion processes used by AI LLMs make it difficult for users to assess the trustworthiness of the information provided. In some cases, AI LLMs present broken or irrelevant links and lack diversity in source selection, raising concerns about the underlying algorithms [31].

The development of retrieval-augmented generation offers a promising strategy for reducing the risk of health misinformation [32]. When implemented through either open or closed models integrated with AI LLMs such as ChatGPT, retrieval-augmented generation enables the delivery of up-to-date, dynamically evolving information while continuously verifying outputs against a linked knowledge base [33]. This enables the creation of high-quality, clear information with improved accuracy and personalization, leading to greater patient satisfaction, particularly in the context of T2DM [34,35].

The average reading age in the United Kingdom is 9 years, suggesting that web pages should ideally be written at a fourth- to fifth-grade reading level [36,37]. Both resource types, however, required an eighth- to ninth-grade reading level for comprehension, which exceeds the national average and may hinder access to health information. Furthermore, this study found no significant correlation between DISCERN scores and the readability of resources. Ideally, higher-quality resources should be easier to read and understand to aid accessibility of health information [38]. Given the high prevalence of T2DM in the United Kingdom, efforts to improve patient education are necessary to address health inequalities [39].

The plethora of online health information available through search engines such as Google, Yahoo, and Bing can be overwhelming for users. Therefore, higher-quality websites, as assessed by the DISCERN instrument, would ideally be more prominently ranked in search results to improve accessibility. Google and Yahoo showed a nonsignificant negative correlation between quality and search position, suggesting that website quality was not a significant predictor of search position. Bing demonstrated a significant positive trend with search rank, indicating that websites appearing higher in the search results tended to have lower DISCERN scores. These findings are consistent with previous research indicating that search engine rankings may not correlate with information quality, warranting further investigation into algorithmic differences and their implications for health information access [40,41].

Limitations

This study should be interpreted within its context. First, although the DISCERN score is a validated and widely used tool for assessing the quality of health information, its application involves a degree of subjective judgment by reviewers. Differences in how individual criteria are interpreted may affect scoring, even when predefined assessment guidelines are used. To minimize this risk, independent assessments were performed by multiple reviewers and interrater reliability was evaluated, demonstrating strong agreement. Nevertheless, the possibility of residual subjective bias cannot be entirely excluded.

Second, the AIGLs were produced using a single standardized prompt across all 4 AI LLMs. Although this ensured uniformity across the leaflets produced, tailoring each AIGL with additional prompts may have yielded higher-quality resources. Additionally, this would replicate users' patterns of using AI LLMs, as multiple prompts are typically used to elicit information from AI.

This study was restricted to text-based resources because assessing videos and interactive tools posed methodological challenges for AI LLMs. However, most of the population now uses social media posts, videos, infographics, and other interactive formats when obtaining health information [42,43]. As content sources vary substantially in quality, readability, and reliability, the findings may not fully reflect the broader digital health information landscape.

The discrepancy in sample sizes between websites (n=75) and AIGL (n=4) remains a limitation due to differences in statistical power. The lack of statistically significant differences observed in this study should not be interpreted as evidence of equivalence between the groups. Given the relatively small sample size, the study may have been underpowered to detect modest but potentially meaningful differences in website quality and reliability metrics. Consequently, the absence of statistical significance may reflect insufficient statistical power rather than a true lack of difference. Future studies incorporating larger sample sizes and a priori power calculations are warranted to confirm these findings and better characterize any underlying differences. This limitation should be considered when interpreting the results and their generalizability.

Finally, given that this study used a cross-sectional design, it does not capture the dynamic nature of websites and AI LLMs. Future studies should be longitudinal to incorporate changes in quality and reliability over time, as these resources are constantly updated and revised.

Conclusions

This study demonstrated that both traditional websites and AI-generated resources offer T2DM information of comparable quality; however, the overall standard remains suboptimal considering the global prevalence and clinical burden of the condition. Additionally, readability levels exceeded recommended thresholds for patient education, raising concerns about accessibility and the potential to exacerbate health inequalities.

Although AI-generated content achieved quality scores similar to those of traditional websites, it did not meet basic reliability criteria as assessed by the JAMA benchmark. The absence of safeguarding measures raises major concerns related to patient safety and governance.

The landscape of medical encounters is evolving, with patients increasingly presenting with preconceived notions based on online health information. Health care professionals should be mindful of this shift during consultations. Prioritizing improvements in the quality, readability, and safeguards of online health information is essential.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Hurtado MD, Vella A. What is type 2 diabetes? Medicine 2019 Jan;47(1):10-15. [doi: 10.1016/j.mpmed.2018.10.010]
2. Type 2 diabetes. International Diabetes Federation. URL: https://idf.org/about-diabetes/types-of-diabetes/type-2/ [accessed 2025-12-18]
3. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet 2014 Jun 7;383(9933):1999-2007. [doi: 10.1016/S0140-6736(14)60613-9] [Medline: 24910231]
4. Type 2 diabetes in adults: management. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng28/chapter/Recommendations#individualised-care [accessed 2023-03-08]
5. Ali MK, Pearson-Stuttard J, Selvin E, Gregg EW. Interpreting global trends in type 2 diabetes complications and mortality. Diabetologia 2022 Jan;65(1):3-13. [doi: 10.1007/s00125-021-05585-2] [Medline: 34837505]
6. Świątoniowska N, Sarzyńska K, Szymańska-Chabowska A, Jankowska-Polańska B. The role of education in type 2 diabetes treatment. Diabetes Res Clin Pract 2019 May;151:237-246. [doi: 10.1016/j.diabres.2019.04.004] [Medline: 31063855]
7. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021 Dec;9(12):1740. [doi: 10.3390/healthcare9121740]
8. Eke R, Jannat KT, Yang XT, Parton JM. Disparity and factors associated with internet health information seeking among US adults living with diabetes mellitus: cross-sectional study. J Med Internet Res 2022 May 30;24(5):e32723. [doi: 10.2196/32723] [Medline: 35635741]
9. Guo F, Zhang X, Lee P. Editorial: digital information for patient education. Front Public Health 2023;11:1211285. [doi: 10.3389/fpubh.2023.1211285] [Medline: 37304109]
10. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
11. Hu D, Guo Y, Zhou Y, Flores L, Zheng K. A systematic review of early evidence on generative AI for drafting responses to patient messages. NPJ Health Syst 2025 Jul 23;2:27. [doi: 10.1038/s44401-025-00032-5]
12. Jeon S, Lee S, Kim EH, et al. Generative AI chatbot for diabetes management: formative 2-part qualitative study using DTalksBot involving patients and clinicians. JMIR Form Res 2025 Nov 12;9:e72553. [doi: 10.2196/72553] [Medline: 41223424]
13. Kim Y, Jeong H, Chen S, et al. Medical hallucinations in foundation models and their impact on healthcare. arXiv. Preprint posted online on Feb 26, 2025. [doi: 10.48550/arXiv.2503.05777]
14. Al-Ak'hali MS, Fageeh HN, Halboub E, Alhajj MN, Ariffin Z. Quality and readability of web-based Arabic health information on periodontal disease. BMC Med Inform Decis Mak 2021 Feb 4;21(1):41. [doi: 10.1186/s12911-021-01413-0] [Medline: 33541345]
15. Beheshti M, Toubal IE, Alaboud K, et al. Evaluating the reliability of ChatGPT for health-related questions: a systematic review. Informatics 2025;12(1):9. [doi: 10.3390/informatics12010009]
16. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
17. Tahir M, Usman M, Muhammad F, et al. Evaluation of quality and readability of online health information on high blood pressure using DISCERN and Flesch-Kincaid tools. Appl Sci 2020;10(9):3214. [doi: 10.3390/app10093214]
18. Wang L, Wan Z, Ni C, et al. Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. J Med Internet Res 2024 Nov 7;26:e22769. [doi: 10.2196/22769] [Medline: 39509695]
19. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of Flesch formula. Educ Health (Abingdon) 2017;30(1):84-88. [doi: 10.4103/1357-6283.210517] [Medline: 28707643]
20. Flesch Kincaid calculator. Good Calculators. URL: https://goodcalculators.com/flesch-kincaid-calculator/ [accessed 2025-12-20]
21. Journal of American Medical Association (JAMA) benchmark criteria. ResearchGate. URL: https://www.researchgate.net/figure/Journal-of-American-Medical-Association-JAMA-benchmark-criteria_tbl1_339976668 [accessed 2025-12-20]
22. Liu J, He J. The decoy dilemma in online medical information evaluation: a comparative study of credibility assessments by LLM and human judges. arXiv. Preprint posted online on Nov 23, 2024. [doi: 10.48550/arXiv.2411.15396]
23. Ahmad MA, Yaramic I, Dutta Roy T. Creating trustworthy LLMs: dealing with hallucinations in healthcare AI. arXiv. Preprint posted online on Sep 26, 2023. [doi: 10.48550/arXiv.2311.01463]

XSL•FO
RenderX

24. AI hallucinates more frequently as it gets more advanced — is there any way to stop it from happening, and should we even try? Live Science. 2025. URL: https://tinyurl.com/ycyt3zv7 [accessed 2026-01-04]
25. OpenAI o3 and o4-mini system card. OpenAI. 2025. URL: https://cdn.openai.com/pdf/2221c875-02dc-4789-800b-e7758f3722c1/o3-and-o4-mini-system-card.pdf [accessed 2026-01-04]
26. Rahman SS, Islam MA, Alam MM, et al. Hallucination to truth: a review of fact-checking and factuality evaluation in large language models. Artif Intell Rev 2026 Jan 19;59:70. [doi: 10.1007/s10462-025-11454-w]
27. Omar M, Agbareia R, Glicksberg BS, Nadkarni GN, Klang E. Benchmarking the confidence of large language models in answering clinical questions: cross-sectional evaluation study. JMIR Med Inform 2025 May 16;13:e66917. [doi: 10.2196/66917] [Medline: 40378406]
28. Lim KH, Kang U, Li X, et al. Susceptibility of large language models to user-driven factors in medical queries. arXiv. Preprint posted online on Mar 26, 2025. [doi: 10.48550/arXiv.2503.22746]
29. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. JNCI Cancer Spectr 2023 Mar 1;7(2):pkad010. [doi: 10.1093/jncics/pkad010] [Medline: 36808255]
30. Ulusoy I, Yılmaz M, Kıvrak A. How efficient is ChatGPT in accessing accurate and quality health-related information? Cureus 2023 Oct;15(10):e46662. [doi: 10.7759/cureus.46662] [Medline: 37942394]
31. Nelson HC, Beauchamp MT, Pace AA. The reliability gap: how traditional search engines outperform artificial intelligence (AI) chatbots in rosacea public health information quality. Cureus 2025 Jun;17(6):e86543. [doi: 10.7759/cureus.86543] [Medline: 40698243]
32. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: a systematic review. PLOS Digit Health 2025 Jun;4(6):e0000877. [doi: 10.1371/journal.pdig.0000877] [Medline: 40498738]
33. Miao J, Thongprayoon C, Suppadungsuk S, Garcia Valencia OA, Cheungpasitporn W. Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications. Medicina (Kaunas) 2024 Mar 8;60(3):445. [doi: 10.3390/medicina60030445] [Medline: 38541171]
34. Baur D, Ansorg J, Heyde CE, Voelker A. Development and evaluation of a retrieval-augmented generation chatbot for orthopedic and trauma surgery patient education: mixed-methods study. JMIR AI 2025 Oct 23;4:e75262. [doi: 10.2196/75262] [Medline: 41134117]
35. Wang D, Liang J, Ye J, et al. Enhancement of the performance of large language models in diabetes education through retrieval-augmented generation: comparative study. J Med Internet Res 2024 Nov 8;26:e58041. [doi: 10.2196/58041] [Medline: 39046096]
36. Reading level for web pages. Surrey County Council. URL: https://www.surreycc.gov.uk/website/writing-guide/reading-levels-for-web-pages [accessed 2026-01-08]
37. The importance of adult literacy levels for elearning. What You Need to Know. URL: https://whatyouneedtoknow.co.uk/adult-literacy-and-elearning/ [accessed 2026-01-08]
38. Manivannan S, Baskaran R, Fisher J, Tasker I, Zaben M. Current status of websites offering information to patients with traumatic brain injury and caregivers: time for reform? World Neurosurg 2021 Sep;153:e419-e427. [doi: 10.1016/j.wneu.2021.06.140] [Medline: 34229103]
39. Tackling Inequality Commission. Diabetes UK. URL: https://www.diabetes.org.uk/for-professionals/improving-care/tackling-inequality-commission [accessed 2026-06-27]
40. Cai HC, King LE, Dwyer JT. Using the Google™ search engine for health information: is there a problem? Case study: supplements for cancer. Curr Dev Nutr 2021 Feb;5(2):nzab002. [doi: 10.1093/cdn/nzab002] [Medline: 33937613]
41. Baskaran R, Gauhar V, Tsaturyan A, et al. Patient information on ureteric stents; online resources versus artificial intelligence generated content: a benchmarking study from EAU endourology. Arab J Urol 2025:1-10. [doi: 10.1080/20905998.2025.2602411]
42. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442. [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
43. Sumayyia MD, Al-Madaney MM, Almousawi FH. Health information on social media. Perceptions, attitudes, and practices of patients and their companions. Saudi Med J 2019 Dec;40(12):1294-1298. [doi: 10.15537/smj.2019.12.24682] [Medline: 31828284]

Abbreviations

AI: artificial intelligence
AI LLM: artificial intelligence large language model
AIGL: artificial intelligence–generated leaflet
FKGL: Flesch-Kincaid Grade Level
JAMA: Journal of the American Medical Association

XSL•FO
RenderX

T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by R Zhao; submitted 19.Feb.2026; peer-reviewed by A Fittler, J Dai; revised version received 02.Jun.2026; accepted 05.Jun.2026; published 15.Jul.2026.

Please cite as:
Srinivasan S, Baskaran R, Mukhopadhyay S, Patel M
Assessing Quality, Readability, and Transparency of Online and AI-Generated Information on Type 2 Diabetes Mellitus: Cross-Sectional Exploratory Study
JMIR Diabetes 2026;11:e93822
URL: https://diabetes.jmir.org/2026/1/e93822
doi:10.2196/93822

© Sripradha Srinivasan, Ravanth Baskaran, Srinjay Mukhopadhyay, Mayank Patel. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 15.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predicting Treatment Failure With Sodium-Glucose Cotransporter-2 Inhibitors in People With Type 2 Diabetes: Novel Artificial Intelligence and Machine Learning Approach

Doyoung Kwak[1], MS; Xi Tan[2], PharmD, PhD; Yuanjie Liang[2], MS; Caroline Swift[2], PhD; Chalak Muhammad[2], MPH, MD; Xu Shi[3], PhD

[1]Texas A&M University, College Station, TX, United States
[2]Novo Nordisk Inc, 800 Scudders Mill Road, Plainsboro, NJ, United States
[3]University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Xi Tan, PharmD, PhD
Novo Nordisk Inc, 800 Scudders Mill Road, Plainsboro, NJ, United States

Abstract

Background: The rate of treatment failure with sodium-glucose cotransporter-2 inhibitors (SGLT2i) is high among individuals with type 2 diabetes (T2D). Accurately predicting SGLT2i treatment failure is important for improving the clinical management of T2D.

Objective: The study aimed to use machine learning (ML) models to identify factors predicting treatment failure with SGLT2i in T2D and to evaluate model performance.

Methods: This retrospective observational cohort study included adults with T2D treated with SGLT2i (2016-2024). The primary outcome was overall treatment failure with SGLT2i during follow-up (≥180 days after SGLT2i initiation). The secondary outcome was subtypes of treatment failure with SGLT2i (treatment discontinuation, failure with action, and inertial failure) or nonfailure, which was defined as not meeting the definition for one of the failure subtypes. Variables potentially associated with treatment failure were assessed during the year before SGLT2i treatment initiation (analysis 1) and the year before SGLT2i treatment failure (analysis 2). Using these variables, ML models—logistic regression (LR), multilayer perceptron (MLP), extreme gradient boosting (XGBoost), and Transformer—were used to identify significant predictors of the outcomes. Model performance metrics (accuracy, area under the curve, precision, recall, and F_1-score) were calculated. Using Shapley Additive Explanations methodology, key features were identified based on their impact on model predictions. LR and Transformer models using key features were further evaluated for their potential to support the development of a risk score for predicting treatment failure with SGLT2i.

Results: Among all individuals in the study (N=62,222), 71% (n=44,156) had treatment failure with SGLT2i. Across subtypes, failure with action (n=23,839, 38.3%) was more common than treatment discontinuation (n=16,449, 26.4%) and inertial failure (n=3868, 6.2%). Model performance was moderate in both analyses. In analysis 1, the accuracy ranged from 0.72 to 0.73 for predicting overall treatment failure and from 0.56 to 0.57 for predicting the subtype of treatment failure. In analysis 2, the accuracy ranged from 0.74 to 0.75 for predicting overall treatment failure and from 0.61 to 0.63 for predicting the subtype of treatment failure. XGBoost, MLP, and Transformer models showed small improvements compared with LR. Using the top 9 key features identified from the Shapley Additive Explanations analysis, the Transformer model performed similarly in accuracy and area under the curve to its counterpart using the full feature set.

Conclusions: Performance across the LR, MLP, XGBoost, and Transformer models was moderate. The advanced ML models performed slightly better than LR. Overall, the results suggest that further model advancements and increased data availability are needed to better predict treatment failure with SGLT2i. The LR coefficients from the key features model may inform the development of a risk score to predict SGLT2i treatment failure. Accurate prediction could inform individualized treatment planning for individuals with T2D.

(JMIR Diabetes 2026;11:e85372) doi:10.2196/85372

KEYWORDS

sodium-glucose cotransporter-2 inhibitors; SGLT2i; type 2 diabetes; artificial intelligence; machine learning; treatment failure

XSL•FO
RenderX

Introduction

Type 2 diabetes (T2D), a chronic and progressive metabolic condition caused by inefficient production or use of insulin and identified by elevated blood glucose levels, affects more than 35 million adults in the United States [1]. Sodium-glucose cotransporter-2 inhibitors (SGLT2i), a commonly prescribed therapy to reduce blood glucose levels in people with T2D, promote the excretion of glucose and reduce its reabsorption into the bloodstream [2]. SGLT2i also have additional benefits beyond glycemic control, including renal and cardiovascular benefits [3-5]. The American Diabetes Association recommends SGLT2i for individuals with T2D and chronic kidney disease (CKD) or kidney damage, as well as those with T2D and heart failure [5].

Despite the benefits of SGLT2i in managing T2D, treatment failure rates are high among individuals in real-world settings [6-8], particularly among those with comorbidities [9]. Treatment failure is a multifaceted problem, and there is currently no consensus on the definition of treatment failure with SGLT2i in people with T2D. Typical components of treatment failure may include discontinuation, switching to other glucose-lowering therapies (GLT), and treatment intensification [9,10]. Treatment failure with SGLT2i may also be characterized by not reaching the desired treatment goal after a certain amount of time or the attenuation of SGLT2i benefits over time [11].

Previous research has identified several factors associated with SGLT2i treatment discontinuation, including older age, specific comorbidities (eg, ischemic heart disease, chronic obstructive pulmonary disease, CKD, and cancer), the type of other GLT used at baseline, adverse effects, lack of efficacy, and financial reasons [7,9,12]. However, the literature lacks a comprehensive assessment of the factors contributing to treatment failure with SGLT2i or its different components, and identifying individuals at higher risk of treatment failure with SGLT2i can be clinically challenging. Identifying individuals for whom SGLT2i treatment is likely to fail has implications for treatment planning and clinical management and, ultimately, the achievement of treatment goals. The prediction of treatment failure with SGLT2i would enable health care professionals to take actions to reduce the likelihood of treatment failure or plan for possible failure based on the individual's unique set of characteristics and circumstances.

Advanced machine learning (ML) models could potentially improve the prediction of treatment failure with SGLT2i in individuals with T2D. By analyzing data on many demographic characteristics, clinical parameters, laboratory values, and treatment histories, advanced ML models may offer advantages over conventional statistical methods for identifying factors and patterns that predict complex, multifaceted outcomes. In contrast to traditional techniques, which often rely on expert-driven a priori feature selection, modern algorithms such as gradient boosting and deep neural networks can automatically screen hundreds of candidate variables, rank their relative importance, and surface nonobvious predictors with minimal manual effort [13,14]. Additionally, by incorporating a broad array of variables, advanced ML models can potentially improve the accuracy and precision of outcome predictions as their flexible architectures capture complex nonlinear relationships and high-order interactions that conventional techniques, such as standard regression frameworks, struggle to represent without extensive handcrafting.

This study aimed to develop and internally validate ML models to predict treatment failure with SGLT2i among people with T2D using linked electronic health records (EHRs) and claims data. Specifically, we aimed to use ML models to predict treatment failure with SGLT2i and its subtypes.

Methods

Study Design and Data Source

This retrospective observational cohort study included adults with T2D treated with SGLT2i. Data were obtained from Optum deidentified Market Clarity Data (Optum Market Clarity), comprising EHRs linked to medical and pharmacy claims from Optum-affiliated payers and additional third-party claims. The database contains records on more than 86 million US people and is considered nationally representative of the US population.

Ethical Considerations

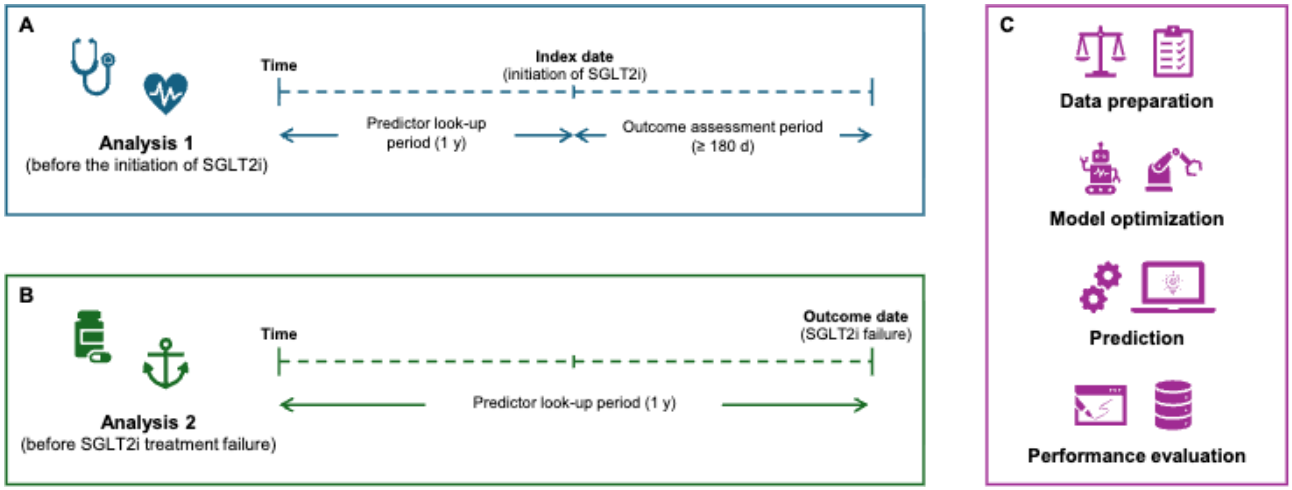
This study was conducted in accordance with the Declaration of Helsinki of 1975 and its subsequent revisions. The study database is Health Insurance Portability and Accountability Act (HIPAA)–compliant and deidentified, and data were analyzed without reidentification of or contact with study participants. Because this study used only previously collected, deidentified data, informed consent from study participants was not necessary. The WCG Institutional Review Board determined this study to be exempt under 45 Code of Federal Regulations (CFR) § 46.104(d)(4) [15].

Data Selection

The study period ran from January 1, 2016, through September 30, 2024. The index date was defined as the date of the first eligible SGLT2i pharmacy claim. The period for identifying eligible individuals for the study ran from December 31, 2016, through April 3, 2024, to include 1 year before the index date and a 180-day outcome assessment period (the interval from the index date to the end of the follow-up). Individuals were followed from the index date to the end of continuous claims enrollment, death, or the end of the study period, whichever occurred first. Two analyses, corresponding to two predictor lookup periods, were conducted to examine predictors of treatment failure with SGLT2i over two periods: (1) the year before SGLT2i treatment initiation (ie, analysis 1; Figure 1); and (2) the year before treatment failure with SGLT2i (ie, analysis 2; Figure 1). It was important to assess both time periods for predictors, as each can offer distinct, yet complementary, information. Analysis 1 focuses on predictors that can inform clinical decision-making before the initiation of treatment, whereas analysis 2 focuses on predictors more proximal to treatment failure, potentially informing clinical decision-making after treatment initiation.

XSL•FO
RenderX

Figure 1. Study schematics. (A) Schematic of analysis 1; (B) schematic of analysis 2; and (C) schematic of data preparation, optimization, prediction, and evaluation steps. SLGT2i: sodium-glucose cotransporter-2 inhibitors; d, days; y, year;



Inclusion and Exclusion Criteria

Individuals were eligible for inclusion in the study if they were newly initiated on an SGLT2i treatment (bexagliflozin, canagliflozin, dapagliflozin, empagliflozin, and ertugliflozin) between December 31, 2016, and April 3, 2024, and had 2 or more diagnoses of T2D on 2 or more distinct days during the study period, with the first T2D diagnosis on or before the index date. Eligible individuals also had continuous claims enrollment in the predictor lookup and outcome assessment periods, were aged 18 years or older on the index date, had 1 or more glycated hemoglobin (HbA_{1c}) value ≥7% during the baseline period or on the index date, and had 1 or more HbA_{1c} value from 180 days after the index date until the end of follow-up. Individuals were excluded if they had 1 or more diagnoses of type 1 diabetes during the study period, had missing data for sex, initiated another new GLT drug class on the index date, used any glucagon-like peptide-1 receptor agonist (GLP-1 RA) obesity medications during the study period, had evidence of pregnancy, or had CKD stage 5 or end-stage kidney disease during the baseline period or on the index date (Figure S1 in Multimedia Appendix 1).

Outcomes

The study outcome was treatment failure with SGLT2i, adopted from our previous work [8], which was defined as the occurrence of any of the following events during the outcome assessment period: initiation of a new GLT class ("failure with action"), discontinuation of SGLT2i ("discontinuation"), or HbA_{1c} not under control ("inertial failure"). Initiation of a new GLT class was defined as any pharmacy claim for a new GLT class (insulin, thiazolidinediones, sulfonylureas, GLP-1 RA, biguanides, meglitinides, alpha-glucosidase inhibitors, or dipeptidyl peptidase 4 inhibitors) during the follow-up period, including either switching to or adding a new GLT class. Discontinuation of SGLT2i was defined as a gap of 90 or more days for any SGLT2i during the follow-up period. Switching within an SGLT2i class was not considered discontinuation. HbA_{1c} was considered uncontrolled if 2 or more consecutive HbA_{1c} laboratory values were 8% or more on distinct days (≥90 days apart) between the index date plus 180 days and the end of the follow-up period. Nonfailure was defined as not meeting the definition for 1 of the failure subtypes. The primary outcome in this study was a binary variable for the occurrence of treatment failure with SGLT2i (treatment failure did or did not occur). The secondary outcome was a multiclass variable for the subtypes of treatment failure—failure with action, treatment discontinuation, or inertial failure. Additional details regarding the definitions of treatment failure with SGLT2i have been previously described [8].

Predictors

Potential predictors of treatment failure with SGLT2i included index year, demographic characteristics (age, sex, race, and geographic region), clinical characteristics (BMI [<25, 25 - 29.9, 30 - 34.9, 35 - 39.9, or ≥40 kg/m^2], kidney function and CKD (stage 1, 2, 3, or 4), time from first observed T2D diagnosis to the index date), laboratory tests (HbA_{1c}, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, very low-density lipoprotein cholesterol, total cholesterol, triglycerides, and C-reactive protein), and vital signs (systolic and diastolic blood pressure measurements; Table S1 in Multimedia Appendix 1). The following baseline comorbidities were also included: atrial fibrillation, ischemic heart disease, deep vein thrombosis or pulmonary embolism, hypertension, dyslipidemia, components of atherosclerotic cardiovascular disease, myocardial infarction, other coronary heart disease, peripheral artery disease, metabolic dysfunction–associated steatohepatitis or metabolic dysfunction–associated steatotic liver disease, obesity, anxiety, depression, asthma, musculoskeletal pain, and osteoarthritis. Other predictors included the Charlson Comorbidity Index adjusted without diabetes [16,17], the Diabetes Complications Severity Index [18,19], tobacco smoking status, GLT use, selected non-GLT treatment use, health care resource utilization (all-cause and T2D-related hospitalizations and emergency department visits), and health care costs (for all-cause and T2D-related hospitalizations and emergency department visits). Detailed information regarding predictor construction can be found in Tables S2-S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Analysis

Model Overview and Performance Evaluation

For analyses 1 and 2, descriptive statistics (mean and SD for continuous variables and counts and percentages for categorical variables) were calculated for demographic characteristics, baseline comorbidities, clinical characteristics, laboratory tests, vital signs, GLT use, selected non-GLT treatment use, health care resource utilization, and health care costs. For both analyses, we evaluated the performance of 4 models—logistic regression (LR), multilayer perceptron (MLP), extreme gradient boosting (XGBoost), and Transformer—to predict treatment failure with SGLT2i. LR, a traditional statistical method known for its simplicity and interpretability, is suitable for binary classification tasks [20]. MLP, XGBoost, and Transformer are advanced ML models. MLP is a feedforward artificial neural network useful for capturing nonlinear relationships through hidden layers [21]. XGBoost is a gradient boosting framework known for its efficiency and accuracy in handling structured data and managing feature interactions [22]. Transformer leverages self-attention mechanisms to process data efficiently and capture complex patterns in diverse datasets [23]. For all analyses, we split the data with 80% as a training set (including model optimization) and 20% as a test set for internal validation. After training and optimization on the training set, we used the test set to evaluate the performance of each model and generate confusion matrices [24]. Model performance was evaluated based on accuracy, receiver operating characteristic area under the curve (ROC AUC), precision, recall, and F_1-score [25,26]. Precision, recall, and F_1-score were computed for the positive class for the binary outcome, and support-weighted averaging was used for the multiclass outcome to account for class imbalance. Analyses were conducted in Python (version 3.10; Python Software Foundation) using *scikit-learn*, XGBoost, and PyTorch, along with supporting libraries such as *NumPy* and *Pandas*.

Data Preparation and Handling of Missing Data

An illustration of the data preparation and processing workflow for each model (LR, MLP, XGBoost, and Transformer) is provided in Figure S2 in Multimedia Appendix 1. In preparing the dataset for the LR model, categorical variables were factorized and transformed into integers based on their unique categories. Numerical variables were normalized using MinMax scaling and centered at the midpoint before model fitting, ensuring all variables were on a comparable scale. Missing values were imputed with 0 on the transformed feature scale, thereby providing a neutral or reference value; because the data were transformed before imputation, an imputed 0 did not indicate an actual clinical measurement of 0. Missing values across all variables (categorical and numerical) were filled with 0 as a consistent approach to minimize their impact on model parameters.

For the MLP model, categorical variables were embedded into vectors, and missing values were filled with 0. Numerical variables were normalized using MinMax scaling. Because a significant number of numerical variables naturally had 0 values, missing numerical values were not filled with 0. To distinguish between actual 0 values and missing numerical values, a masking strategy was used to flag missing data, with each numerical vector concatenated with a binary mask to indicate missingness. This approach ensured that data abnormalities were managed without distortion and facilitated the MLP model's ability to integrate both categorical embeddings and numerical inputs efficiently.

For the XGBoost model, categorical variables in the dataset were converted to a string format to ensure consistent handling, using XGBoost's built-in method of setting "unknown" values to missing. XGBoost inherently handles missing data by learning default directions for nodes in decision trees and assigning optimal paths for instances with missing values during training. This approach simplifies data preparation and maintains performance without requiring imputation of missing values.

For the Transformer model, categorical variables were processed through embeddings, converting them into dense vector representations, with any missing values filled with 0. As with the MLP model, numerical variables were normalized using MinMax scaling, and missing values were handled by creating a mask to indicate the presence of missing data. This approach to missingness also facilitated the Transformer's ability to use information on both missing and nonmissing values effectively. By integrating positional encoding, the Transformer model encapsulates categorical and numerical features and their complex interactions, enhancing its predictive capabilities on tabular data.

In routinely collected EHRs or claims data, missing data are often informative because they are driven by care processes rather than random measurement error [27]. We did not use multiple imputation to impute missing data for any model because multiple imputation relies on strong assumptions (eg, missing-at-random) and a well-specified imputation model. Applying multiple imputation indiscriminately when missingness is structurally driven by routine care may introduce bias or clinically implausible values [28].

Parameterization

For the MLP and Transformer models, hyperparameter optimization, including tuning learning rate, batch size, number of epochs, and architecture, was conducted using a representative subset of 3000 individuals from the training dataset. This subset of 3000 individuals was randomly sampled from the training set to reduce the substantial computational burden associated with hyperparameter optimization for advanced ML models. The optimized hyperparameters were then applied to train the full dataset [29]. For the XGBoost model, optimization was performed through 5-fold cross-validation on the training set, evaluating accuracy and F_1-scores across combinations of hyperparameters (eg, gamma, maximum depth, number of estimators, and learning rate). The optimal set of hyperparameters was selected based on cross-validation results, and the final model was retrained on the full training data before evaluation on the test set. The Appendix provides details on the optimization strategy (Table S2 in Multimedia Appendix 1), optimized hyperparameters (Table S3 in Multimedia Appendix 1), and final training iterations or epochs on the full training split (Table S4 in

XSL•FO
RenderX

Multimedia Appendix 1) for each model. The same hyperparameter sets and optimization strategies were applied in analyses 1 and 2.

Feature Importance

Feature importance scores were derived to identify the contribution of individual variables in predicting SLGT2i treatment failure. Model-specific techniques were used to assess feature importance for each model. In the LR model, we compared the absolute values of coefficients and sorted them by magnitude to determine each feature's relative influence. In the XGBoost model, the model's internal function was used for importance calculation, focusing on metrics such as gain, cover, and frequency. "Gain" indicates the contribution of a feature to the model's predictive capabilities, "cover" assesses the number of observations affected by splits on the feature, and "frequency" counts the occurrences of a feature in decision trees, thereby revealing the feature's significance. In the Transformer model, a Shapley Additive Explanations (SHAP) analysis was conducted to interpret feature importance [30]. SHAP attributes the contribution of each feature according to its impact on model predictions. To enhance clinical utility, defined as the extent to which a predictive model informs clinical decision-making and actions to improve patient care, and support individualized treatment strategies, we identified key features based on their importance values in the best models. We ranked candidate predictors by their mean absolute SHAP value (global importance) and then selected features with clinical utility for further analysis to inform the development of a risk score or algorithm to predict treatment failure of SGLT2i. Upon conducting model comparisons, it was observed that the Transformer model performed slightly better than the other models while maintaining stability. Therefore, we focused on the Transformer model for key feature identification.

Results

Baseline Characteristics

A total of 62,222 individuals with T2D who initiated treatment with SGLT2i during the study period were included in the analyses (Tables S5 and S6 in Multimedia Appendix 1). More than half of the sample (n=35,487, 57%) was male, the mean (SD) age was 62.7 (12.0) years, and most individuals were White (n=44,774, 72%). Most individuals had commercial insurance (n=27,378, 44%) or Medicare (n=29,882, 48%). The most common comorbidities were hyperlipidemia (n=50,851, 81.7%), hypertension (n=50,681, 81.5%), and obesity (n=28,374, 45.6%; Table S5 in Multimedia Appendix 1). Overall, 71% (n=44,156) of individuals in the study experienced treatment failure with SGLT2i. When subtypes were assessed, failure with action (n=23,839, 38.3%) was more common than discontinuation (n=16,449, 26.4%) and inertial failure (n=3868, 6.2%; Table S5 in Multimedia Appendix 1).

Features Analysis and Model Evaluation Results

In both analyses 1 and 2, using the full dataset, model performance was moderate overall (Table 1). Compared with the LR model, the Transformer, XGBoost, and MLP models showed small improvements (Table 1). In analysis 1, accuracy across models ranged from 0.72 to 0.73 for overall failure or not (binary outcome) and from 0.56 to 0.57 for the subtypes of failure (multiclass outcome). ROC AUC across models ranged from 0.69 to 0.70 for overall failure and from 0.63 to 0.64 for the failure subtypes (Figure 2). In analysis 2, accuracy across models ranged from 0.74 to 0.75 for overall failure and from 0.61 to 0.63 for the failure subtypes. ROC AUC ranged from 0.74 to 0.75 for overall failure and from 0.72 to 0.75 for the failure subtypes (Figure 3). Similar patterns were observed for precision, recall, and F_1-score estimates in both analyses 1 and 2. Notably, models developed for analysis 2 consistently outperformed those from analysis 1, particularly in the subtypes of the failure prediction task (Table 1).

XSL•FO
RenderX

Table . Performance metrics for models used to predict treatment failure with SGLT2i[a] among people with type 2 diabetes.

Model	Accuracy	ROC AUC[b]	Precision	Recall	F_1-score
Analysis 1[c]					
Overall failure (failure vs not)					
LR[d]	0.72	0.69	0.73	0.96	0.83
MLP[e]	0.72	0.69	0.74	0.92	0.82
XGBoost[f]	0.73	0.70	0.74	0.95	0.83
Transformer	0.73	0.70	0.73	0.96	0.83
Subtype of failure (discontinuation, failure with action, and inertial failure) or nonfailure					
LR	0.56	0.63	0.53	0.56	0.50
MLP	0.56	0.63	0.50	0.56	0.46
XGBoost	0.57	0.64	0.53	0.57	0.50
Transformer	0.57	0.63	0.56	0.57	0.49
Analysis 2[c]					
Overall failure (failure vs not)					
LR	0.74	0.74	0.76	0.93	0.84
MLP	0.74	0.74	0.76	0.93	0.84
XGBoost	0.74	0.74	0.77	0.92	0.84
Transformer	0.75	0.75	0.77	0.91	0.83
Subtype of failure (discontinuation, failure with action, and inertial failure) or nonfailure					
LR	0.61	0.73	0.60	0.61	0.59
MLP	0.61	0.72	0.60	0.61	0.59
XGBoost	0.63	0.75	0.62	0.63	0.61
Transformer	0.61	0.73	0.61	0.61	0.60

[a]SGLT2i: sodium-glucose cotransporter-2 inhibitor.

[b]ROC AUC: receiver operating characteristic area under the curve.

[c]Analyses were conducted to examine predictors of treatment failure with SGLT2i over two periods: (1) the year before SGLT2i treatment initiation (ie, analysis 1) and (2) the year before treatment failure with SGLT2i (ie, analysis 2).

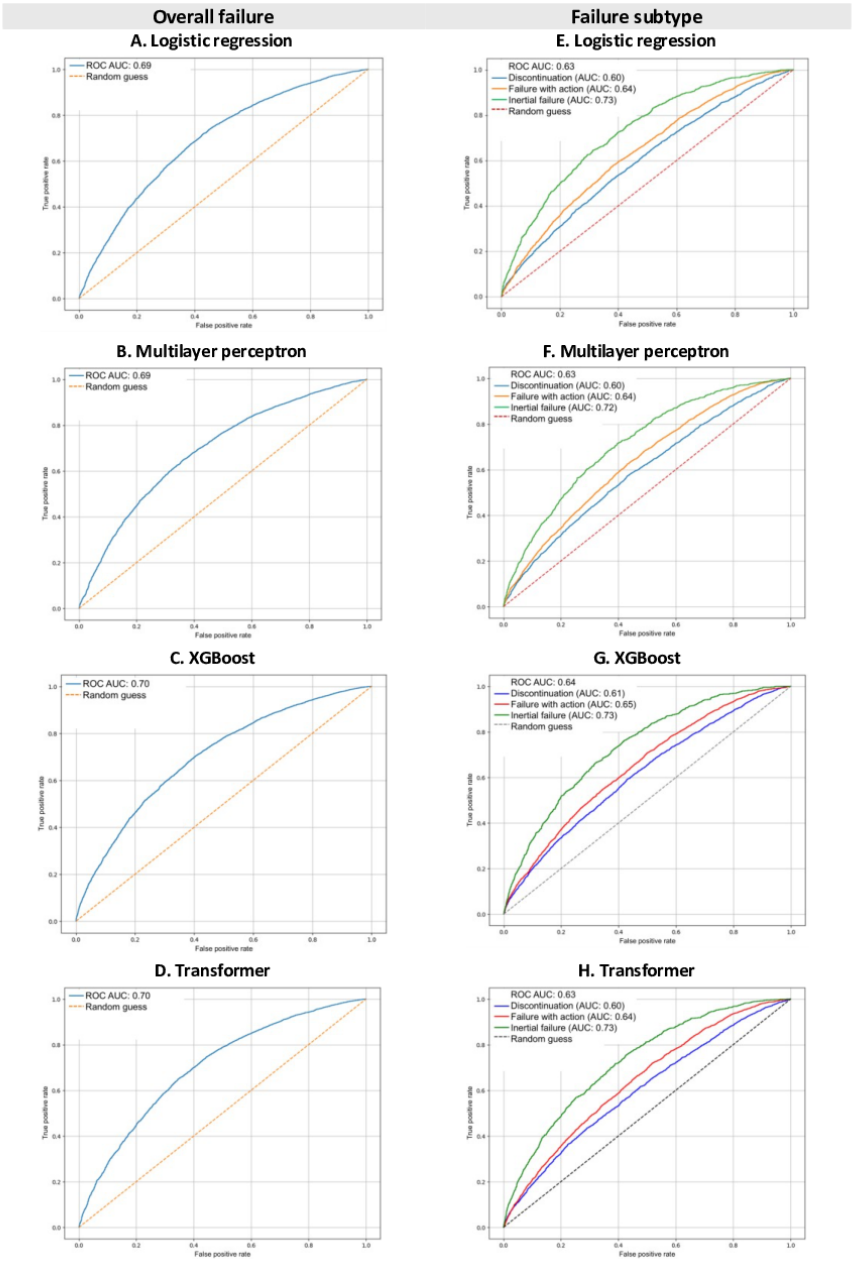
[d]LR: logistic regression.

[e]MLP: multilayer perceptron.

[f]XGBoost: extreme gradient boosting.

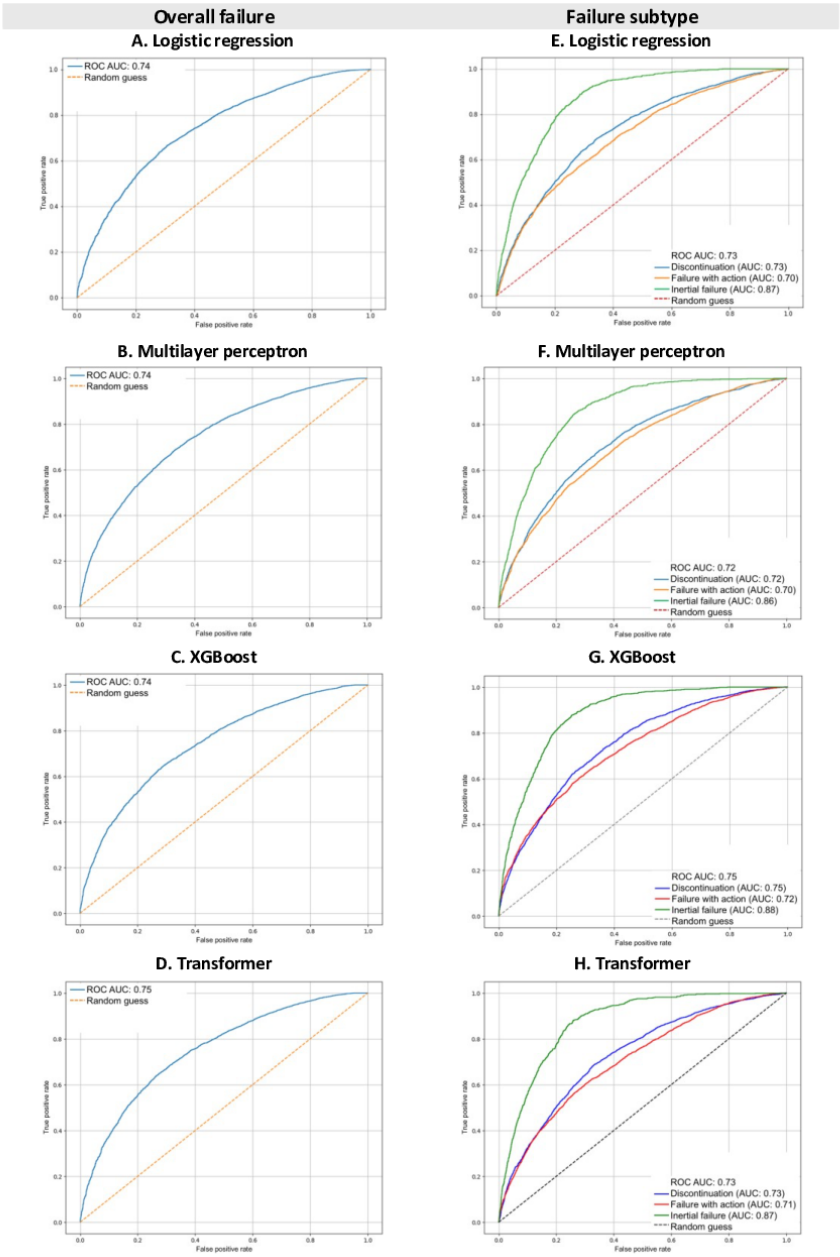
XSL•FO
RenderX

Figure 2. Receiver operating characteristic (ROC) curves (analysis 1). AUC: area under the curve; XGBoost: extreme gradient boosting. Models used to predict overall failure: (A) logistic regression; (B) multilayer perceptron; (C) XGBoost; (D) Transformer. Models used to predict failure subtype: (E) logistic regression; (F) multilayer perceptron; (G) XGBoost; (H) Transformer.



XSL•FO
RenderX

Figure 3. Receiver operating characteristic (ROC) curves (analysis 2). AUC: area under the curve; XGBoost: extreme gradient boosting. Models used to predict overall failure: (A) logistic regression; (B) multilayer perceptron; (C) XGBoost; (D) Transformer. Models used to predict failure subtype: (E) logistic regression; (F) multilayer perceptron; (G) XGBoost; (H) Transformer.

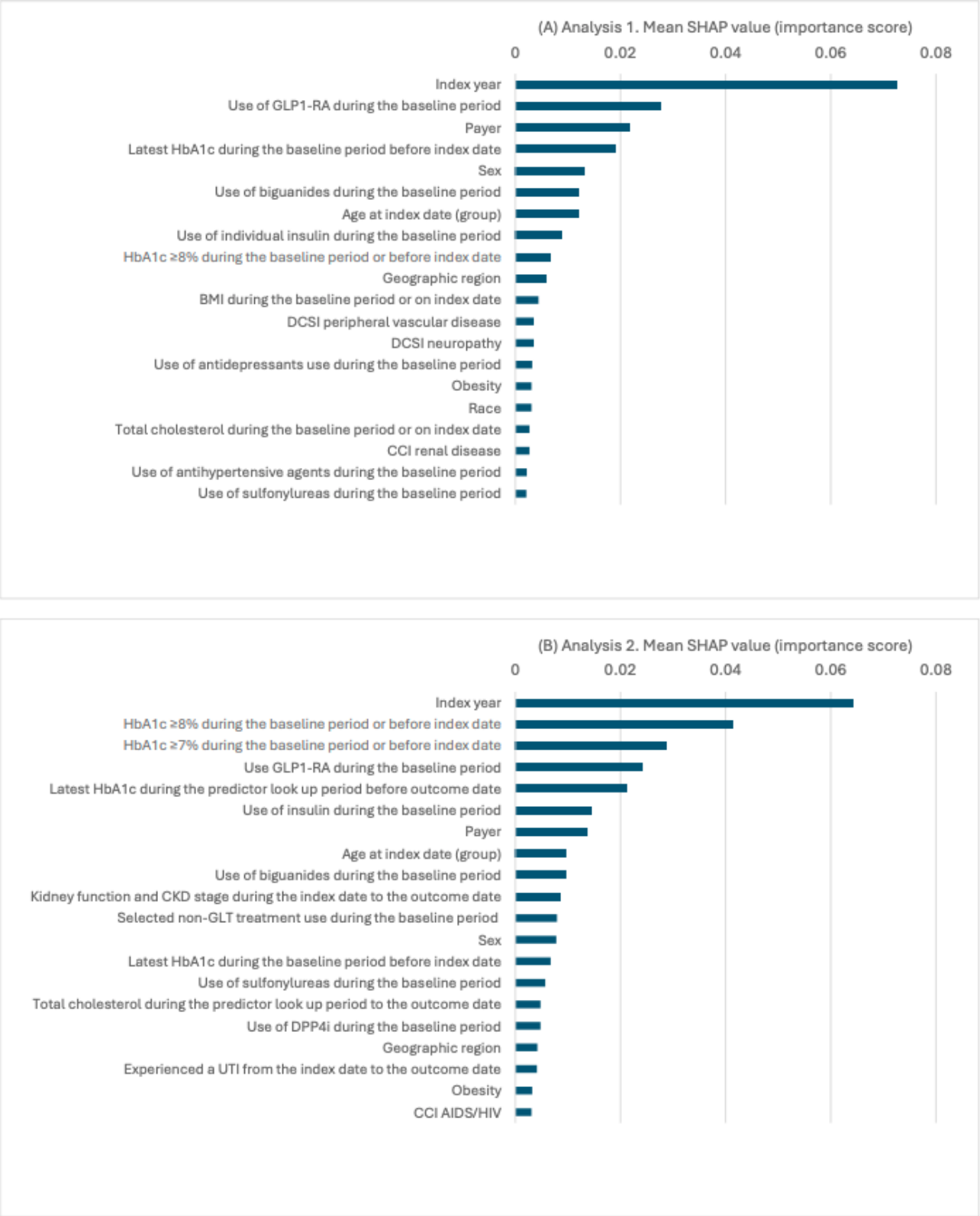


Feature Importance and Key Features Analysis

The SHAP analysis based on the Transformer model uncovered consistent feature importance patterns across analyses 1 and 2. Index year, HbA_{1c}, and use of GLP-1 RA at baseline exhibited the highest importance scores in both analyses. Additionally, comparative examination of mean SHAP values and visualizations for the remaining individual features revealed similarities across the 2 analyses, with nuanced variations in the relative magnitude of these features. Analysis 2 also identified features after SGLT2i initiation, such as having experienced a urinary tract infection (Figure 4).

XSL•FO
RenderX

Figure 4. Transformer model in analysis 1 and analysis 2: feature importance bar plot from Shapley Additive Explanations (SHAP) analysis. (A) Analysis 1. Mean SHAP Value (Importance Score); (B) Analysis 2. Mean SHAP Value (Importance Score). CCI: Charlson Comorbidity Index; CKD: chronic kidney disease; DCSI: Diabetes Complication Severity Index; DPP4i: dipeptidyl peptidase 4 inhibitor; GLP-1 RA: glucagon-like peptide-1 receptor agonist; GLT: glucose-lowering therapy; HbA_{1c}: glycated hemoglobin; UTI: urinary tract infection.



We identified key features for further analysis to support the development of a risk score or algorithm to predict treatment failure with SGLT2i. Because the index year, regarded as a proxy for follow-up time, may not be useful in clinical practice given the study period of this particular analysis and its potential limited utility in clinical decision-making, we excluded it from further analysis of key features. We tested both Transformer and LR models to evaluate their effectiveness using the remaining 9 top features (Figure 4): HbA_{1c}≥8% during the baseline period or before the index date, use of GLP-1 RA

XSL•FO
RenderX

during the baseline period; HbA_{1c}≥7% during the baseline period or before the index date, use of insulin during the baseline period; and latest HbA_{1c} during the predictor lookup period before the outcome date, payer type, use of biguanides during the baseline period, age at index date (group), and kidney function and CKD stage during the index date to the outcome date. The number of key features was not determined by a statistical SHAP threshold; rather, it was chosen pragmatically to support clinical utility by limiting the number of required inputs and enabling a parsimonious model that could be operationalized as a simple risk score. Using the 9 selected key features, the Transformer model had slightly higher accuracy (0.75 vs 0.74) and recall (0.98 vs 0.97) and the same ROC AUC (0.70), precision (0.75), and F_1-score (0.85) compared with the LR model (Table 2). The coefficients from the LR model with the 9 key features are presented in Table 3 and may inform further development of a risk score or algorithm to conveniently predict treatment failure with SGLT2i in routine clinical practice.

Table . Performance from the logistic regression and Transformer models with 9 key features.

Model	Accuracy	ROC AUC[a]	Precision	Recall	F_1-score
LR model with 9 key features	0.74	0.70	0.75	0.97	0.85
Transformer model with 9 key features	0.75	0.70	0.75	0.98	0.85

[a]ROC AUC: receiver operating characteristic area under the curve.

Table . The coefficients from the logistic regression (LR) model with the 9 key features.

Key feature	LR coefficient (log odds)
HbA_{1c}[a] ≥8% during the baseline period or before index date	0.89
Use of GLP-1 RA[b] during the baseline period	–0.50
HbA_{1c} ≥7% during the baseline period or before index date	–0.40
Use of insulin during the baseline period	–0.33
Latest HbA_{1c} during the predictor lookup period before outcome date	0.16
Payer	–0.15
Use of biguanides during the baseline period	0.14
Age at index date (group)	–0.02
Kidney function and CKD[c] stage from the index date to the outcome date	0.02
Constant (intercept)	–0.49

[a]HbA_{1c}: glycated hemoglobin.

[b]GLP-1 RA: glucagon-like peptide-1 receptor agonist.

[c]CKD: chronic kidney disease.

Discussion

This real-world, observational study showed that treatment failure with SGLT2i is prevalent among people with T2D in the United States, which is consistent with our previous work [8]. We found that ML models produce modest performance improvements relative to traditional LR for predicting treatment failure with SGLT2i. Furthermore, we identified 9 predictors that may be particularly influential in predicting treatment failure with SGLT2i, including those readily available in EHR for individuals with T2D, such as HbA_{1c} levels, use of GLT, and kidney function. This study's strengths included using a nationally representative US database with a large population and examining predictors in 2 reference periods—the year before treatment initiation with SGLT2i and the year before treatment failure with SGLT2i. This study also used advanced ML approaches to capture complex, nonlinear relationships and automatically identify important predictors from high-dimensional clinical data. Feature importance analysis was conducted to better explain the prediction models and enhance their utility in clinical practice.

This study provides important information for health care professionals on the high treatment failure rate with SGLT2i among people with T2D in real-world settings and the key predictors that may increase the risk of treatment failure. Overall, across analyses 1 and 2, model performance was moderate [31,32]. Although improvement in model performance is desirable, the findings suggest that advanced ML models, including MLP, XGBoost, and Transformer, can learn relevant patterns from the data. Furthermore, in analyses 1 and 2, these advanced ML models showed modest improvements in performance metrics over traditional LR, indicating a slightly better ability to capture nuanced patterns. The superior model performance in analysis 2 was likely due to postindex data

XSL•FO
RenderX

providing information more proximal to treatment failure with SGLT2i, enabling more precise predictions. Having information post-SGLT2i initiation (index date), especially on potential SGLT2i treatment-related adverse events, may help boost predictive accuracy.

Building on the analyses with the full feature set, we examined the LR and Transformer models using a reduced set of 9 features selected based on importance scores. Both models demonstrated comparable performance, with scores reaching moderate levels overall. Notably, these models performed similarly to those using the full feature set, underscoring the potential utility of parsimonious models incorporating influential predictors. In the context of T2D and SGLT2i treatment, feature ranking facilitates the identification of the most influential predictors for determining treatment failure with SGLT2i. From a clinical perspective, ML and strategic feature selection could empower health care professionals by helping them identify individuals at the highest risk of treatment failure and other treatment-related outcomes, such as medication nonadherence, for further treatment planning or intervention [33-35].

Although ML models hold promise for clinical prediction, data quality issues, the need for large datasets, costs, and implementation challenges in real-world settings impede the uptake of ML-based prediction models in clinical practice [36]. LR is commonly used to predict clinical outcomes, including treatment failure, in clinical practice settings due to its familiarity among health care professionals and its ability to handle binary outcomes. Coefficients from LR models offer insight into the relationship between a set of predictors and the likelihood of treatment failure. Coefficients from the LR model used in our study could inform the development of a future prediction tool to estimate the likelihood of treatment failure with SGLT2i among individuals with T2D. Until such a prediction tool is available, the key features identified in this study (eg, HbA_{1c}≥7%, payer type, or use of insulin, biguanides, or GLP-1 RA before SGLT2i initiation) could be of value for clinical decision-making to reduce treatment failure among people with T2D.

Future research is needed to refine and validate ML models across populations and settings to increase the applicability of ML-based predictive results to clinical practice in T2D while addressing deficiencies in data quality and quantity [37]. Specifically, there is a need to increase the breadth of predictors to encompass health behaviors, family history, genetic information, social determinants of health, patient-reported outcomes, and comprehensive laboratory data. Concurrently, data quality must also be addressed to improve its accuracy and completeness, thereby enhancing the efficiency and utility of predictive analytics. Furthermore, although ML methodologies, including deep learning, have demonstrated efficacy in medical imaging, diagnosis, and early detection of disease or complications, their performance in predicting clinical outcomes is suboptimal. There is a need to develop ML models that can utilize complex, high-dimensional data to produce highly accurate, clinically relevant information [38-40]. Given the modest performance gains observed, LR remains a strong baseline for clinical deployment due to its superior interpretability. Our results also indicate that the current data and composite end point likely cap performance across models, while advanced architectures may provide meaningful value with richer, longitudinal, or multimodal inputs for tasks closer to the failure event, as suggested by the consistently better performance in analysis 2. Finally, the 9-feature model serves as a pragmatic bridge to clinical use by retaining comparable performance with far greater simplicity.

This study had some limitations. Although ML models can identify features strongly correlated with treatment failure with SGLT2i, these correlations do not necessarily imply causation. This is an inherent limitation of observational study design. Second, the datasets used to build and validate the models comprised administrative claims-linked EHR data, which were not collected for research purposes. These data may include potential inaccuracies in diagnostic coding, measurement error, and incomplete information on variables used in the predictive models. Different strategies for handling missing data may enable some models to use missingness patterns more directly than others, potentially influencing model performance. However, these strategies reflect inherent differences in model architecture and standard practice and are unlikely to have driven model performance because differences between advanced models and LR were small overall. Although explicit masking could confer an advantage for deep learning models, we observed only modest improvements. Given the large sample size and the substantial computational burden associated with training advanced models—particularly XGBoost and the Transformer—we used a fixed 80%/20% training/testing split for final model evaluation. Although performance metrics based on a single internal split may not fully capture variability across data partitions, this approach enabled consistent comparisons of performance. Regarding the design of predictors, we aimed at clinical interpretability, with conditions and drug utilization over a period of time summarized. A more complex design incorporating the longitudinal nature of predictors (ie, the timing and sequence) or more free-form predictors (eg, unlabeled diagnosis codes) may improve the predictive accuracy of some of the models but at a loss of clinical interpretation. Furthermore, the datasets used in this study did not contain data on certain factors that could be important for predicting treatment failure with SGLT2i, such as more granular clinical characteristics, behavioral factors, and social determinants of health. Furthermore, although this study's focus was a parsimonious, clinically implementable single-risk score, subtype-specific models could further disentangle underlying mechanisms of treatment failure with SGLT2i, which could be an important avenue for future research. Finally, this study focused on the glycemic control indication for T2D and did not include other indications for SGLT2i (eg, cardiovascular disease).

This study found a high prevalence of treatment failure with SGLT2i among people with T2D. In predicting overall treatment failure and subtype, model performance was moderate. Models based on the examination of predictors in the year before treatment failure with SGLT2i performed better than models based on the examination of predictors in the year before SGLT2i initiation. The results suggest that further advancements in ML models and additional data are needed to enhance

XSL•FO
RenderX

ML-based prediction of treatment failure with SGLT2i. Feature importance analysis may support the development of a risk score or algorithm to inform more timely individualized treatment planning for individuals with T2D.

Acknowledgments

The authors thank Yiwen Cao for providing data management support. Generative artificial intelligence was not used in the preparation of this manuscript. The authors thank Benjamin Doty and Rebecca Hahn of KJT Group, Inc, for providing medical writing support, in accordance with Good Publication Practice and International Committee of Medical Journal Editors guidelines.

Funding

This work was funded by Novo Nordisk Inc.

Data Availability

The datasets generated and/or analyzed in this study are not publicly available because they were commercially licensed from the data vendor. Restrictions apply to the availability of these data, which were used under license for this study.

Authors' Contributions

Conceptualization: XT, YL, CS, CM, DK
Data curation: XT, DK
Formal analysis: DK
Investigation: XT, YL, CS, CM, DK, XS
Methodology: XT, YL, CS, CM, DK, XS
Project administration: XT
Visualization: XT, DK
Writing – original draft: XT, YL, CS, CM, DK, XS
Writing – review and editing: XT, YL, CS, CM, DK, XS

Conflicts of Interest

XT, YL, CS, and CM are employees of Novo Nordisk Inc. YL, CS, and CM are shareholders of Novo Nordisk A/S. DK is a former employee of Novo Nordisk Inc. XS is a consultant for Novo Nordisk Inc.

Multimedia Appendix 1
Supplemental material.
[DOCX File, 314 KB - diabetes_v11i1e85372_app1.docx]

References

1. National Diabetes Statistics Report. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/php/data-research [accessed 2025-03-28]
2. Klen J, Dolžan V. Treatment response to SGLT2 inhibitors: from clinical characteristics to genetic variations. Int J Mol Sci 2021 Sep 10;22(18):9800. [doi: 10.3390/ijms22189800] [Medline: 34575958]
3. Padda IS, Mahtani AU, Parmar M. Sodium-glucose transport protein 2 (SGLT2) inhibitors. In: StatPearls: StatPearls Publishing; 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK576405/ [accessed 2026-02-04]
4. de Boer IH, Khunti K, Sadusky T, et al. Diabetes management in chronic kidney disease: a consensus report by the American Diabetes Association (ADA) and Kidney Disease: Improving Global Outcomes (KDIGO). Diabetes Care 2022 Dec 1;45(12):3075-3090. [doi: 10.2337/dci22-0027] [Medline: 36189689]
5. American Diabetes Association Professional Practice Committee. 9. Pharmacologic approaches to glycemic treatment: standards of care in diabetes—2025. Diabetes Care 2025 Jan 1;48(Supplement_1):S181-S206. [doi: 10.2337/dc25-S009] [Medline: 38078590]
6. Fadini GP, Tentolouris N, Caballero Mateos I, Bellido Castañeda V, Morales Portillo C. A multinational real-world study on the clinical characteristics of patients with type 2 diabetes initiating dapagliflozin in Southern Europe. Diabetes Ther 2020 Feb;11(2):423-436. [doi: 10.1007/s13300-019-00744-6] [Medline: 31863344]
7. Malik ME, Falkentoft AC, Jensen J, et al. Discontinuation and reinitiation of SGLT-2 inhibitors and GLP-1R agonists in patients with type 2 diabetes: a nationwide study from 2013 to 2021. Lancet Reg Health Eur 2023 Jun;29:100617. [doi: 10.1016/j.lanepe.2023.100617] [Medline: 37265783]
8. Tan X, Cao Y, Wang Y, et al. Prevalence, treatment patterns, burden, and factors related to treatment failure with sodium-glucose cotransporter 2 inhibitor in adults with type 2 diabetes in the United States. Diabetes Metab Syndr 2025 Jul;19(7):103281. [doi: 10.1016/j.dsx.2025.103281] [Medline: 40819425]

XSL•FO
RenderX

9. Kovesdy C, Schmedt N, Folkerts K, et al. Predictors of cardio-kidney complications and treatment failure in patients with chronic kidney disease and type 2 diabetes treated with SGLT2 inhibitors. BMC Med 2022 Jan 10;20(1):2. [doi: 10.1186/s12916-021-02191-2] [Medline: 35000594]
10. Mocarski M, Yeaw J, Divino V, et al. Slow titration and delayed intensification of basal insulin among patients with type 2 diabetes. J Manag Care Spec Pharm 2018 Apr;24(4):390-400. [doi: 10.18553/jmcp.2017.17218] [Medline: 29406841]
11. Singhal M, Tan H, Coleman CI, Han M, Nguyen C, Ingham M. Effectiveness, treatment durability, and treatment costs of canagliflozin and glucagon-like peptide-1 receptor agonists in patients with type 2 diabetes in the USA. BMJ Open Diabetes Res Care 2019;7(1):e000704. [doi: 10.1136/bmjdrc-2019-000704] [Medline: 31798890]
12. Montvida O, Shaw J, Atherton JJ, Stringer F, Paul SK. Long-term trends in antidiabetes drug usage in the U.S.: real-world evidence in patients newly diagnosed with type 2 diabetes. Diabetes Care 2018 Jan;41(1):69-78. [doi: 10.2337/dc17-1414] [Medline: 29109299]
13. Alsahaf A, Petkov N, Shenoy V, Azzopardi G. A framework for feature selection through boosting. Expert Syst Appl 2022 Jan;187:115895. [doi: 10.1016/j.eswa.2021.115895]
14. Fitz S, Romero P. Neural networks and deep learning: a paradigm shift in information processing, machine learning, and artificial intelligence. In: Rau R, Wardrop R, Zingales L, editors. The Palgrave Handbook of Technological Finance: Springer International Publishing; 2021:589-654. [doi: 10.1007/978-3-030-65117-6_22]
15. Code of federal regulations (annual edition). govInfo. URL: https://www.govinfo.gov/app/collection/cfr [accessed 2026-04-16]
16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
17. Glasheen WP, Cordier T, Gumpina R, Haugh G, Davis J, Renda A. Charlson Comorbidity Index: ICD-9 update and ICD-10 translation. Am Health Drug Benefits 2019;12(4):188-197. [Medline: 31428236]
18. Young BA, Lin E, Von Korff M, et al. Diabetes complications severity index and risk of mortality, hospitalization, and healthcare utilization. Am J Manag Care 2008 Jan;14(1):15-23. [Medline: 18197741]
19. Glasheen WP, Renda A, Dong Y. Diabetes complications severity index (DCSI)-update and ICD-10 translation. J Diabetes Complications 2017 Jun;31(6):1007-1013. [doi: 10.1016/j.jdiacomp.2017.02.018] [Medline: 28416120]
20. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. J Educ Res 2002 Sep;96(1):3-14. [doi: 10.1080/00220670209598786]
21. Bishop CM. Neural Networks for Pattern Recognition: Oxford University Press; 1995. URL: https://academic.oup.com/book/52815 [accessed 2026-04-04]
22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016 URL: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf [accessed 2026-04-29]
23. Vaswani A, Shazeer N, Parmar N. Attention is all you need. Presented at: Adv Neural Inf Process Syst (NIPS 2017); Dec 4-9, 2017 URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf [accessed 2026-04-29]
24. Ting KM. Confusion matrix. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning: Springer; 2010:209. [doi: 10.1007/978-0-387-30164-8_157]
25. Manning CD, Raghavan P, Schutz H. Chapter 8. evaluation in information retrieval. In: Manning CD, Raghavan P, Schutz H, editors. Introduction to Information Retrieval: Cambridge University Press; 2008:139-161 URL: https://www.cambridge.org/highereducation/books/introduction-to-information-retrieval/669D108D20F556C5C30957D63B5AB65C#overview [accessed 2026-04-04]
26. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
27. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. Diagn Progn Res 2020;4(1):8. [doi: 10.1186/s41512-020-00077-0] [Medline: 32699824]
28. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338(jun29 1):b2393. [doi: 10.1136/bmj.b2393] [Medline: 19564179]
29. Rochette T. Ian Goodfellow-deep learning-2016. Tom Rochette. URL: https://blog.tomrochette.com/agi/books/ian-goodfellow-deep-learning#11-4-selecting-hyperparameters [accessed 2025-09-04]
30. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Presented at: Advances in neural information processing systems; Dec 4-9, 2017. [doi: 10.48550/arXiv.1705.07874]
31. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316. [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]
32. Çorbacıoğlu Ş, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: a guide to interpreting the area under the curve value. Turk J Emerg Med 2023;23(4):195-198. [doi: 10.4103/tjem.tjem_182_23] [Medline: 38024184]

XSL•FO
RenderX

33. Kanyongo W, Ezugwu AE. Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives. Inform Med Unlocked 2023;38:101232. [doi: 10.1016/j.imu.2023.101232]
34. Wu XW, Yang HB, Yuan R, Long EW, Tong RS. Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms. BMJ Open Diabetes Res Care 2020 Mar;8(1):e001055. [doi: 10.1136/bmjdrc-2019-001055] [Medline: 32156739]
35. Lo-Ciganic WH, Donohue JM, Thorpe JM, et al. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care 2015 Aug;53(8):720-728. [doi: 10.1097/MLR.0000000000000394] [Medline: 26147866]
36. Hofer IS, Burns M, Kendale S, Wanderer JP. Realistically integrating machine learning into clinical practice: a road map of opportunities, challenges, and a potential future. Anesth Analg 2020 May;130(5):1115-1118. [doi: 10.1213/ANE.0000000000004575] [Medline: 32287118]
37. Scott IA, Cook D, Coiera EW, Richards B. Machine learning in clinical practice: prospects and pitfalls. Med J Aust 2019 Sep;211(5):203-205.e1. [doi: 10.5694/mja2.50294] [Medline: 31389031]
38. Zhu T, Li K, Herrero P, Georgiou P. Deep learning for diabetes: a systematic review. IEEE J Biomed Health Inform 2021 Jul;25(7):2744-2757. [doi: 10.1109/JBHI.2020.3040225] [Medline: 33232247]
39. Khalid S, Kim H, Kim HS. Recent trends in diabetes mellitus diagnosis: an in-depth review of artificial intelligence-based techniques. Diabetes Res Clin Pract 2025 Jun;224:112221. [doi: 10.1016/j.diabres.2025.112221] [Medline: 40328407]
40. Kwak D, Liang Y, Shi X, Tan X. Comparing machine learning and advanced methods with traditional methods to generate weights in inverse probability of treatment weighting: the INFORM study. Pragmat Obs Res 2024;15:173-183. [doi: 10.2147/POR.S466505] [Medline: 39386162]

Abbreviations

CKD: chronic kidney disease
EHR: electronic health record
GLP-1 RA: glucagon-like peptide-1 receptor agonist
GLT: glucose-lowering therapy
HbA_{1c}: glycated hemoglobin
HIPAA: Health Insurance Portability and Accountability Act
LR: logistic regression
ML: machine learning
MLP: multilayer perceptron
ROC AUC: receiver operating characteristic area under the curve
SGLT2i: sodium-glucose cotransporter-2 inhibitor
SHAP: Shapley Additive Explanations
T2D: type 2 diabetes
XGBoost: extreme gradient boosting

Edited by I Steenstra, J Lu; submitted 09.Oct.2025; peer-reviewed by J Ma, X Yu; revised version received 13.Feb.2026; accepted 26.Feb.2026; published 20.May.2026.

Please cite as:
Kwak D, Tan X, Liang Y, Swift C, Muhammad C, Shi X
Predicting Treatment Failure With Sodium-Glucose Cotransporter-2 Inhibitors in People With Type 2 Diabetes: Novel Artificial Intelligence and Machine Learning Approach
JMIR Diabetes 2026;11:e85372
URL: https://diabetes.jmir.org/2026/1/e85372
doi:10.2196/85372

© Doyoung Kwak, Xi Tan, Yuanjie Liang, Caroline Swift, Chalak Muhammad, Xu Shi. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 20.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Antidiabetic Drug Associations With Heart Failure Outcomes: Real-World Evidence Study Using Electronic Health Records

Elzbieta Jodlowska-Siewert[1], MD, MS; Yunhui Chen[2], BS; Sinian Zhang[1], BS; Jia Li[3], PhD; Robert Dellavalle[4], MD, MSPH, PhD; Rui Zhang[3], PhD; Jue Hou[1], PhD

[1]Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, 2221 University Ave SE, Suite 200, Minneapolis, MN, United States

[2]School of Statistics, College of Liberal Arts, University of Minnesota, Minneapolis, MN, United States

[3]Department of Surgery, Division of Computational Health Sciences, Medical School, University of Minnesota, Minneapolis, MN, United States

[4]Department of Dermatology, Medical School, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Jue Hou, PhD
Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, 2221 University Ave SE, Suite 200, Minneapolis, MN, United States

Abstract

Background: Patients with type 2 diabetes mellitus (T2D) have a higher risk of cardiovascular disease, including heart failure (HF), leading to health care burden including hospitalization and mortality. Among multiple T2D therapies, there are inadequate head-to-head comparisons of their effects on HF in the real-world patient population.

Objective: This study aims to compare the time-to-HF among patients treated with different T2D drugs following metformin.

Methods: We conducted a retrospective data analysis on electronic health records of 5000 patients with T2D. The inclusion criteria were previous treatment with metformin and initiation of glucagon-like peptide-1 receptor agonists (GLP1 RAs), dipeptidyl peptidase-4 inhibitors (DPP4i), sulfonylureas, or insulin. We grouped patients by the mechanism of their subsequent therapies and focused on 2 pairs of comparisons classified by insulin resistance: sulfonylureas versus insulin (increased resistance) and GLP1 RA versus DPP4i (decreased resistance). The outcomes were 5-year HF status and the HF-free survival time, which was verified manually by examining clinical notes. We applied doubly robust causal estimation and accounted for confounding by adjusting for coded and natural language processing electronic health record features identified through medical knowledge networks.

Results: The study included 939 patients, of whom 204 (21.7%) received insulin, 482 (51.3%) received sulfonylureas, 90 (9.6%) received GLP1 RA, and 163 (17.4%) received DPP4i. Patients in the sulfonylureas group had a significantly higher 5-year HF-free survival compared to the insulin group (survival ratio of insulin/sulfonylureas 0.902, 95% CI 0.840 - 0.976; P=.01). There was no significant difference between the DPP4i versus GLP1 RA group in 5-year HF-free survival (survival ratio of GLP1 RA/DPP4i was 0.953, 95% CI 0.849 - 1.067; P=.40). For the occurrence of a HF-related hospitalization within 5 years, there were no significant differences between the sulfonylureas and insulin groups (risk difference 0.057, 95% CI –0.011 to 0.132; P=.11), and between the GLP1 RA and DPP4i groups (risk difference 0.010, 95% CI –0.096 to 0.129).

Conclusions: We evaluated real-world evidence on 2 head-to-head comparisons of second-line T2D therapies on 5-year HF outcomes. Patients on sulfonylureas were associated with lower 5-year HF risks than those treated with insulin when measured by risk ratio, but no significant difference was detected when measured by the risk difference. Limitations of this study included potentially inadequate adjustment of confounding in the observational study and a limited sample size with validated HF outcomes.

(JMIR Diabetes 2026;11:e85083) doi:10.2196/85083

KEYWORDS

comparative effectiveness; natural language processing; glucagon-like peptide-1 receptor agonists; dipeptidyl peptidase-4 inhibitors; sulfonylureas; insulin

Introduction

Diabetes mellitus is a chronic disease caused by impaired glucose metabolism [1]. In 2021, more than 38 million US adults (14.7%) had diabetes, with an additional 8.7 million (3.4%) people who remained undiagnosed. It was estimated that 97.6 million people lived with prediabetes [2], with type 2 diabetes (T2D) accounting for 90% of all cases [3]. Patients with T2D have a 2 to 4 times higher risk of cardiovascular diseases (CVDs), and on average develop CVDs almost 15 years sooner

XSL•FO
RenderX

than nondiabetic individuals [4,5], leading to morbidity and mortality [6]. Heart failure (HF) is one of the most expensive CVDs in patients with T2D [7], because HF exacerbations are a common cause of hospitalization [8-10]. Five-year HF-related hospitalization rates in individuals with T2D vary between 3% and 44% [11-13]. Furthermore, HF has the highest 30-day rate of rehospitalizations, with up to 25% of readmissions caused by HF exacerbation [14]. Therefore, control of HF-related hospitalizations is an important goal for T2D management [9,14].

Several medications have been developed and approved for treating T2D, which are often grouped by their mechanisms of action, including glucagon-like peptide-1 receptor agonists (GLP1 RA), dipeptidyl peptidase-4 inhibitors (DPP4i), and sodium-glucose cotransporter 2 inhibitors (SGLT2i). Due to its availability, affordability, effectiveness, and side effect profile, metformin has long been a typical first-line treatment up until recently and still remains highly recommended [15]. Some of these medications do not solely target short-term glycemic control but also improve long-term outcomes such as obesity and lower risks of CVDs [16,17]. As traditional treatment options, insulin and sulfonylureas are still widely used for their availability, affordability, and effectiveness in glycemic control when other treatments fall short [18]. Among the treatments, SGLT2i has demonstrated cardiovascular benefits on HF outcomes in clinical trials and is now used to treat HF [19-21]. There is evidence that some GLP1 RA drugs improve symptoms of HF in patients with obesity with preserved ejection fraction [22], while some DPP4i drugs may increase HF-related hospitalization rates [23], but there is no proven class effect in all patients with HF for any other T2D medication [24]. Electronic health records (EHRs) provide a data source for real-world evidence [25-28] when direct evidence from clinical trials is absent [29]. As part of routine care, EHRs document medical notes and codes for administrative and billing purposes, which also contain necessary information for clinical variables, such as diagnoses, prescriptions, procedures, laboratory tests, and vital signs. While filtering for study-related information from large databases can be challenging [30], recent development of knowledge networks extracted from longitudinal EHRs now provides a means for effective and efficient confounding variable selection [31,32]. For a key clinical variable, such as the HF outcomes, studies have reported discrepancies between disease onset status and apparent EHR codes matched by description [33-36]. Hence, standardized manual chart review is commonly used to validate results [37].

In this study, we aim to use EHRs to evaluate the impact of GLP1 RA, DPP4i, insulin, and sulfonylureas on HF in patients with T2D for whom SGLT2i are unavailable, contraindicated, or unsuccessful in achieving glycemic targets. Considering the stark heterogeneity in patient characteristics associated with HF profile, for example, access to care and insulin resistance [7,38], we focused on the comparison within traditional treatments inducing insulin resistance (insulin vs sulfonylureas) and newer treatments reducing insulin resistance (GLP1 RA vs DPP4i). To characterize potential confounding, we identified EHR features (codes or terms from notes) relevant to T2D through a knowledge network [32] and adjusted for their counts in pretreatment windows by causal estimation [39,40].

Methods

Ethical Considerations

The study was approved by the Institutional Review Board at the University of Minnesota with waiver for informed consent (STUDY00018213). Patients who opted out of the use of EHRs for research were excluded from the study, with a waiver for informed consent. The analysis was conducted with limited data hosted in an HIPPA compliant server to minimize the risks to participants' privacy and confidentiality.

Study Data

We extracted the data from the EHRs of 201,212 patients who received care at M Health Fairview and had T2D diagnosis codes identified by the *International Classification of Diseases version 9* and *version 10* under the PheWAS catalog PheCode 250.2 [41] (Table S1 in Multimedia Appendix 1). Because some patients with T2D diagnosis codes might not have T2D due to mistakes in documentation or nondiagnosis use of the code for billing, we deployed a multimodal automated phenotyping (MAP) algorithm to further filter for patients likely to have T2D [42]. MAP is a validated, unsupervised algorithm that classifies disease status based on counts of diagnosis codes and mentions in notes for the target disease adjusted by the total number of health care encounters [42,43]. A clinically trained annotator reviewed the disease status of 50 randomly selected patients to validate the accuracy of diagnosis codes and MAP predictions. We chose a 95% specificity cutoff for the MAP T2D filter and defined the phenotyped T2D-positive patients as those with MAP predictions greater than the cutoff. We randomly subsampled 5000 patients from the phenotyped T2D positive patients. We extracted extensive EHR data for this cohort, including all EHR codes (diagnosis, procedure order and result, medication, and laboratory test result) and narrative notes. We deployed a natural language processing (NLP) pipeline [44] to recognize the medical terms in notes compiled in the Unified Medical Language System [45].

Study Design

We considered 4 groups of medications as interventions, including insulin; sulfonylureas (chlorpropamide, glyburide, glipizide, glimepiride, and tolbutamide); DPP4i (linagliptin, alogliptin, saxagliptin, and sitagliptin); and GLP1 RA (dulaglutide, lixisenatide, albiglutide, exenatide, liraglutide, and semaglutide; Figure S1 in Multimedia Appendix 1). We did not include the following medications approved to treat T2D as interventions of interest in the comparisons: SGLT2i (empagliflozin, canagliflozin, dapagliflozin, ertugliflozin, bexagliflozin, and sotagliflozin) for their established benefits of reducing HF risks [46,47]; and thiazolidinediones for their contraindication in patients with HF [48]. Patients who received SGLT2i prior to treatments of interest were still included in the study, with their past SGLT2i treatment used as a baseline variable. The index date (time 0) of the study was defined as the initiation of the first treatment among the list of interest. Inclusion criteria were patients aged older than 18 years, prior

XSL•FO
RenderX

metformin treatment (Figure S2 in Multimedia Appendix 1), established care >365 days prior to index date, and maintained care >30 days after index date (Figure S3 in Multimedia Appendix 1). We identified potential treatment with metformin and subsequent medications of interest by their RxNorm medication codes [49]. The inclusion criterion of previous metformin treatment was based on the long-time role as a typical first-line treatment throughout the observation window in our data [15]. Besides, requiring at least 1 previous treatment may reduce the data leakage issue of baseline T2D characteristics from past treatments at another health institute. To filter out potential spurious codes [50], we required a pair of medication codes for the same ingredient to occur 30 to 180 days apart, reflecting the revolving cycles of established T2D treatments. For treatment identified by this rule, we considered the date of the first medication code in the qualifying pair as the treatment initiation date. Eligible patients were assigned to insulin, sulfonylureas, DPP4i, or GLP1 RA groups based on their treatment after being on metformin. We also extracted the treatment data in the first 5 years of the follow-up to summarize the continuation of metformin or switching to treatments of a different class (Table S2 in Multimedia Appendix 1). We defined established and maintained care as having at least 1 medical encounter occurring more than 365 days prior to the index date and more than 30 days after the index date.

The primary HF outcome in the study was HF-related hospitalization or mortality up to 5 years from the index date. Mortality information was extracted from death records in the EHRs. A clinically trained annotator reviewed the medical charts of eligible patients from the index date to 5 years after to determine the presence of HF-related hospitalization and the corresponding date. We considered the participants censored at the date of their last medical encounter in EHRs.

To measure confounding, we compiled an extensive list of baseline variables. We extracted sex; age at baseline; race; ethnicity; baseline BMI; rural or urban residence; duration of T2D (from the date of first T2D diagnosis code to the index date); duration of metformin treatment (from date of identified metformin initiation to the index date); past HF history determined by the occurrence of HF diagnosis code; past treatment of SGLT2i; and baseline laboratory tests including glycated hemoglobin, low-density lipoprotein, high-density lipoprotein, and total cholesterol. Laboratory tests and BMI values were retrieved within a year before baseline and marked as missing data if no measurement was taken during the year; if more than 1 measurement was available, we selected the one closest to baseline. Missing test values were imputed by the mean of complete cases. In addition to these expert-selected variables, we adjusted for a broad list of EHR features related to T2D or CVD, according to the knowledge network Online Narrative and Codified Feature Search Engine [32]. The full list consisted of 251 diagnosis PheCodes [41], 55 laboratory test Logical Observation Identifier Names and Codes [51], 43 medication RXNORM codes [49], 13 procedure Clinical Classification Software codes [52], and 386 NLP concept unique identifiers [45]. These are all common ontologies rolled up from raw EHR codes [31]. To better characterize the baseline CVD risks, we additionally summarized baseline CVD medications according to their mechanisms (Figure S4 in Multimedia Appendix 1). We summarized these EHR features by counting in 2 temporal windows: more than 1 year prior to the index date for medical history and within 1 year prior to the index date for recent medical conditions (Figure S3 in Multimedia Appendix 1). We filtered out the variables that occurred in fewer than 5% (47/939) of the patients.

Statistical Analysis

For each comparison, we estimated 2 treatment effect parameters: risk differences in 5-year HF-related hospitalization or death and the ratio in 5-year non-HF survival rate. We applied doubly robust estimation methods for both treatment effect parameters that optimally integrate outcome regression and inverse propensity score weighting to achieve robustness to model specifications [40,53]. To select confounders from the large number of candidates, we used the adaptive least absolute shrinkage and selection operator to estimate the propensity score model, logistic outcome regression model for risk difference, and additive hazards outcome regression model for no-HF survival ratio [54]. We accounted for censoring in risk difference analysis by inverse probability of censoring weighting, assuming independent censoring. We used bootstrap to calculate the SD of the estimates and normal approximation to calculate the 95% CIs. Our primary analysis is an intent-to-treat style that ignores the postindex treatment switching. As a sensitivity analysis, we performed a per-protocol style analysis that excluded all patients who switched to another treatment class.

All analyses were conducted in R (version 4.3.1; R Foundation for Statistical Computing). *P* values less than .05 were considered significant.

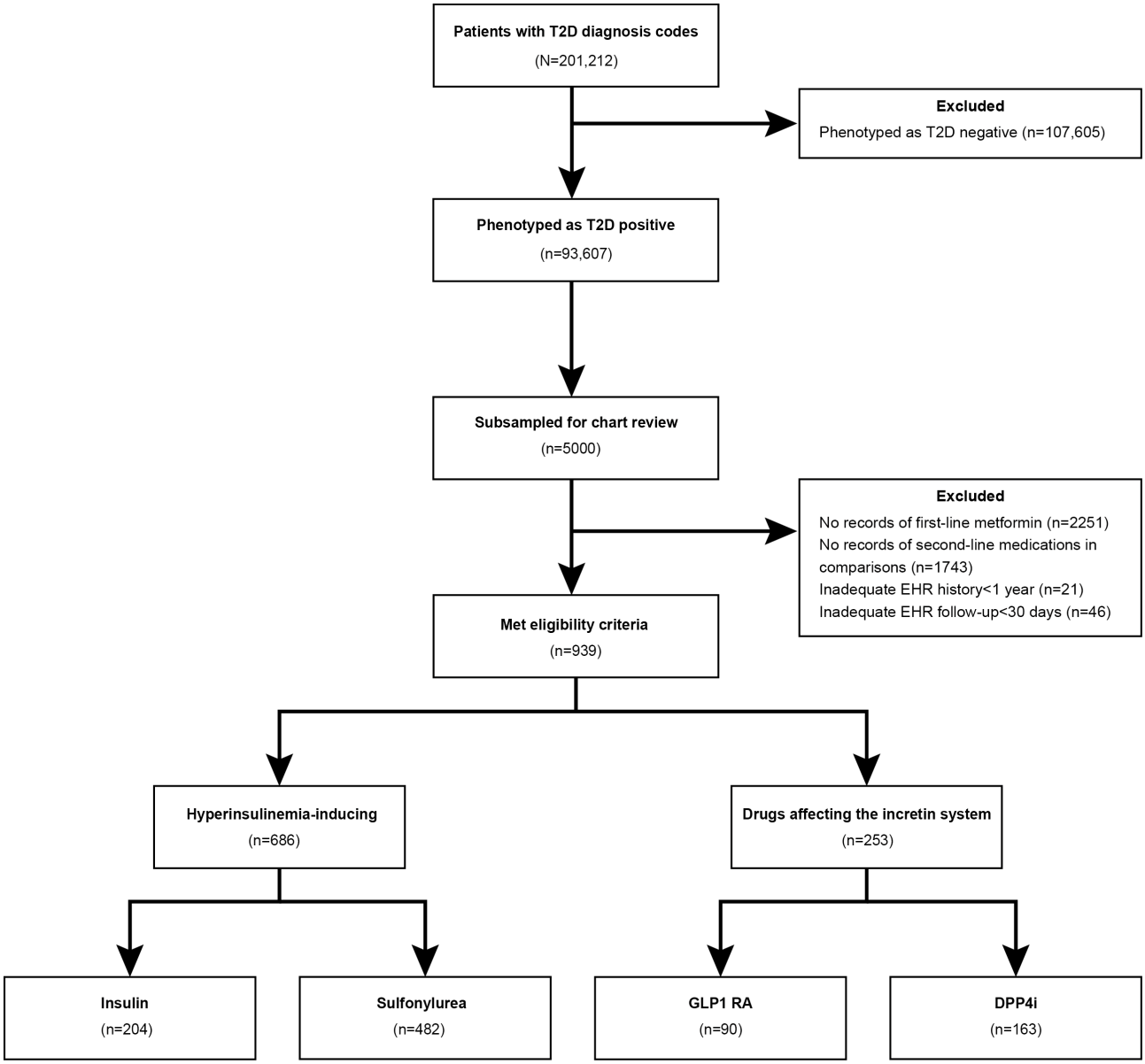
Results

Study Cohort

A total of 93,607 patients were phenotyped as T2D positive using the MAP algorithm, from which we subsampled 5000 patients for validation of their data through chart review. A subset of 939 patients satisfied all eligibility criteria for comparisons in HF outcomes, including 204 who received insulin, 482 who received sulfonylurea, 90 who received GLP1 RA, and 163 who received DPP4i (Figure 1). Among the 939 patients included in the comparisons, 411 (43.8%) were female and 528 (56.2%) were male; median baseline age was 59.4 (IQR 52.1 - 67.9) years; 922 (98.2%) were urban residents and 17 (1.8%) were rural residents; 9 (1%) were American Indian or Alaska Native, 65 (6.9%) were Asian, 81 (8.6%) were Black, 1 (0.1%) was Pacific Islander, 780 (83.1%) were White, and 3 (0.3%) were of more than one race; 19 (2%) were Hispanic; median duration of diabetes was 3.9 (IQR 1.4 - 7.0) years; median duration of metformin use was 0.7 (IQR 0.0 - 2.6) years; baseline hemoglobin A_{1c} was 8.2% (IQR 7.1% - 9.2%); 49 (5.2%) had preexisting HF (Table 1); and 6 (0.6%) had used SGLT2i previously (Table S2 in Multimedia Appendix 1). From the manual chart review, we identified HF-related hospitalization for 132 (14.1%) patients during the follow-up.

XSL•FO
RenderX

Figure 1. Construction of study cohort. DPP4i: dipeptidyl peptidase-4 inhibitors; EHR: electronic health record; GLP1 RA: glucagon-like peptide-1 receptor agonists; T2D: type 2 diabetes.



XSL•FO
RenderX

Table . Characteristics of eligible patients included in the study (N=939).

Variable or level	Summary
Age (y), median (IQR)	59.4 (52.1 - 67.9)
Sex, n (%)	
Female	411 (43.8)
Male	528 (56.2)
Residence, n (%)	
Urban	922 (98.2)
Rural	17 (1.8)
Race, n (%)	
American Indian or Alaska Native	9 (1)
Asian	65 (6.9)
Black	81 (8.6)
Native Hawaiian or other Pacific Islander	1 (0.1)
White	780 (83.1)
More than 1 race	3 (0.3)
Ethnicity, n (%)	
Hispanic	19 (2)
Not Hispanic	920 (98)
Diabetes duration (y), median (IQR)	3.9 (1.4 - 7.0)
Duration of metformin treatment (y), median (IQR)	0.7 (0.0 - 2.6)
Hemoglobin A_{1c} level (%), median (IQR)	8.2 (7.1 - 9.2)
Preexisting heart failure, n (%)	49 (5.2)

Confounding Adjustments

Besides standard confounding factors, such as age, BMI, and preexisting HF, our analysis identified and adjusted for additional confounding factors among the variables compiled form knowledge networks, including a recent diagnosis of edema, recent prescriptions of SGLT2i, past prescriptions of perflutren, and a contrast agent used for heart ultrasound imaging indicative for baseline cardiovascular conditions (Figure 2). After adjusting for the propensity score, all demographic and clinical variables were balanced between the treatment groups in comparison (Tables S3 and S4 in Multimedia Appendix 1).

XSL•FO
RenderX

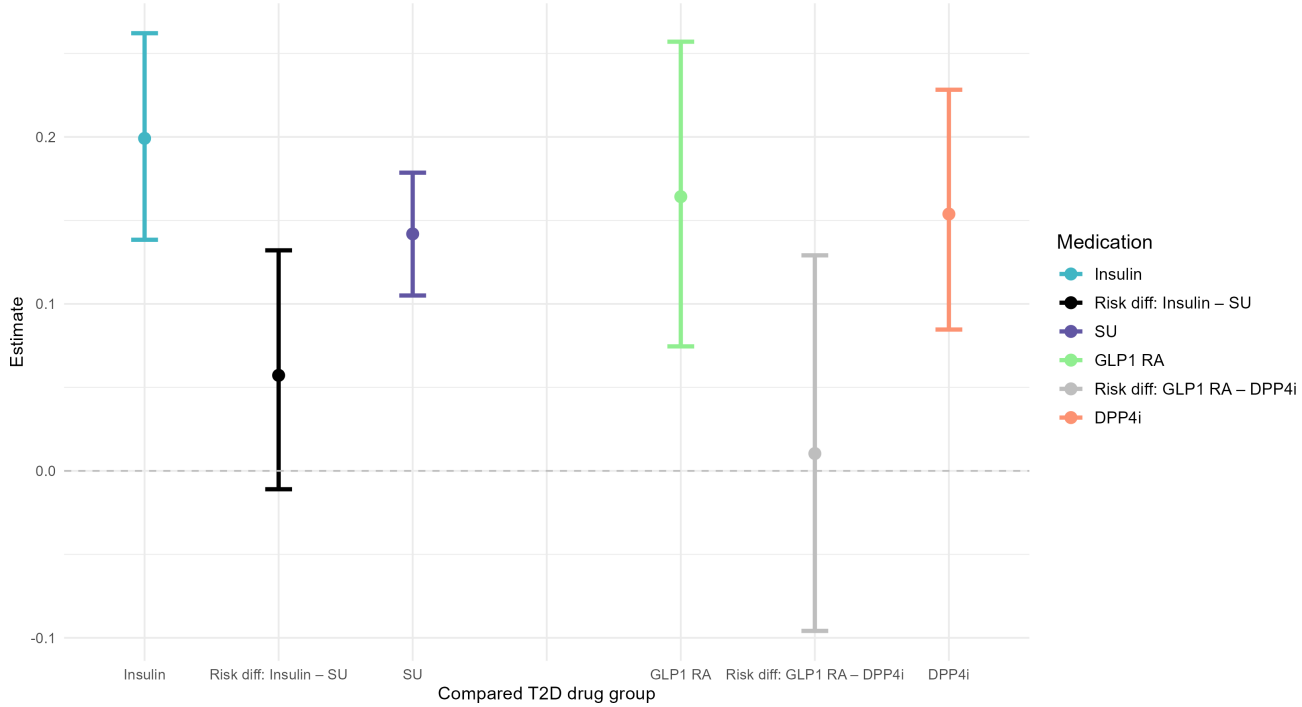
Figure 2. Adjusted confounding in analysis. EHR: electronic health record; SGLT2i: sodium-glucose cotransporter 2 inhibitors.

XSL•FO
RenderX

Insulin Versus Sulfonylureas

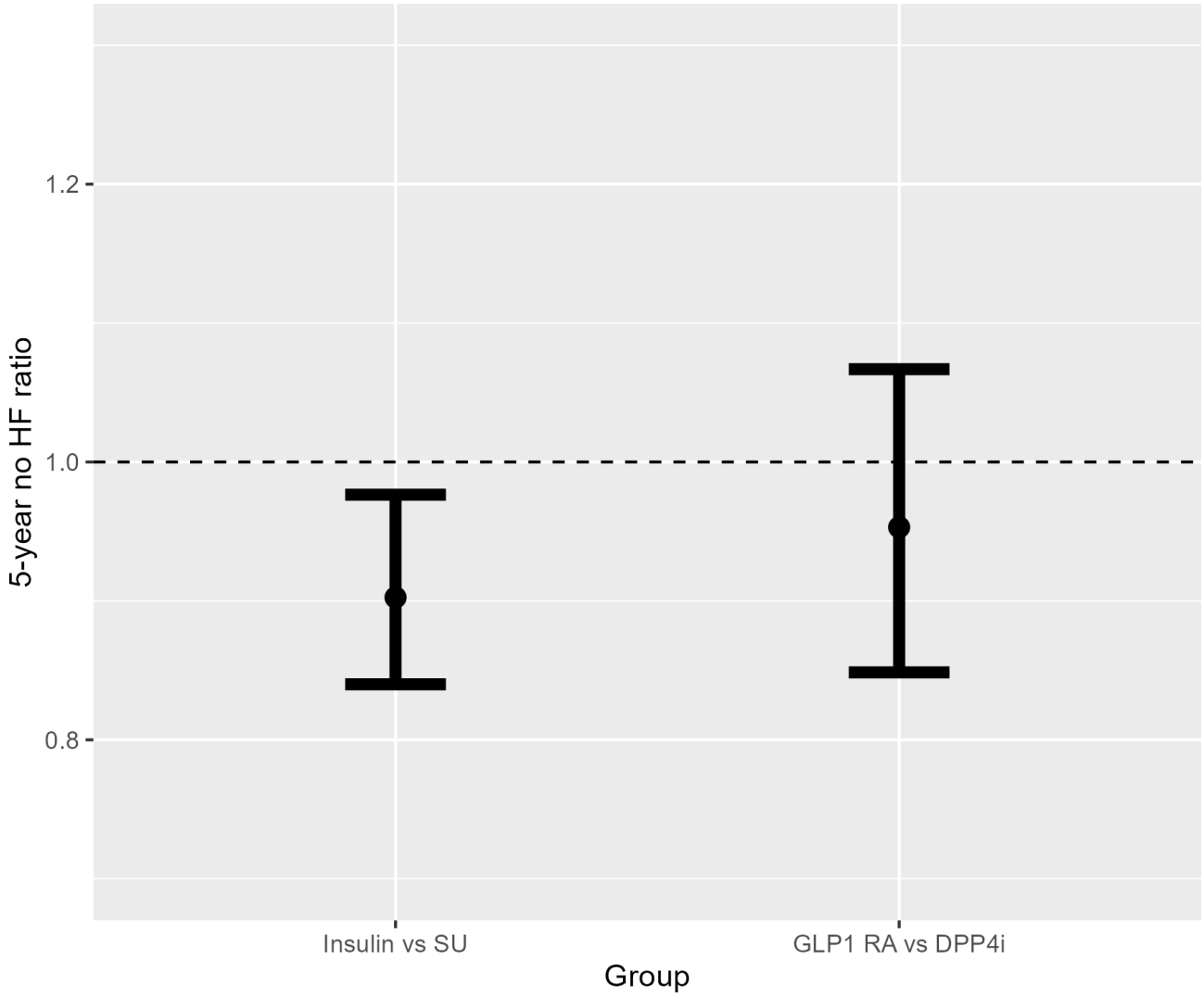
After adjusting for confounding and censoring, we estimated that 0.199 (95% CI 0.138 - 0.262) of patients who received insulin had a HF-related medical encounter within 5 years, compared to 0.142 (95% CI 0.105 - 0.179) of patients who received sulfonylureas, with a risk difference 0.057 (95% CI –0.010 to 0.132; P=.11; Figure 3). We estimated that the 5-year HF-free survival ratio for insulin versus sulfonylureas was 0.902 (95% CI 0.840 - 0.976; P=.01; Figure 4).

Figure 3. Analysis results on risk differences. DPP4i: dipeptidyl peptidase-4 inhibitors; GLP1 RA: glucagon-like peptide-1 receptor agonists; SU: sulfonylureas; T2D: type 2 diabetes.



XSL•FO
RenderX

Figure 4. Analysis results on HF-free survival ratios. DPP4i: dipeptidyl peptidase-4 inhibitors; GLP1 RA: glucagon-like peptide-1 receptor agonists; HF: heart failure; SU: sulfonylureas.



GLP1 RA Versus DPP4i

After adjusting for confounding and censoring, we estimated that 0.164 (95% CI 0.075 - 0.257) of patients who received GLP1 RA agonists had a HF-related medical encounter within 5 years, compared to 0.154 (95% CI 0.085 - 0.228) of patients who received DPP4i, with a risk difference 0.010 (95% CI –0.096 to 0.129; P=.85; Figure 3). We estimated that the 5-year HF-free survival ratio for GLP1 RA versus DPP4i was 0.953 (95% CI 0.849 - 1.067; P=.40; Figure 4).

Secondary and Sensitivity Analyses

We present the results of the per-protocol analysis and a comparison of treatments designed to increase versus reduce insulin resistance in Multimedia Appendix 1. The exclusion of patients who switched treatment resulted in a decrease in the estimated HF-related hospitalization for GLP1 RA, DPP4i, and sulfonylurea, resulting in a significant difference for sulfonylurea versus insulin (risk difference 0.103, 95% CI 0.004 - 0.197; P=.04). We did not observe significant differences for all other analyses.

Discussion

Principal Findings

In this observational study, we found that while patients receiving sulfonylureas had better HF-free survival when compared to those receiving insulin by risk ratio, the risk difference was not significant. Compared to the risk difference analysis, the HF-free survival ratio analysis additionally used the timing of HF-related health care encounters within 5 years, which potentially increased its testing power. Our comparisons of GLP1 RA versus DPP4i failed to detect significant differences in 5-year HF outcomes with the current sample size.

Our findings concerning insulin and sulfonylureas are consistent with previous studies, which examined all CVDs, and not HF specifically. Insulin was reported to be associated with a higher rate of CVD events [55] compared to sulfonylureas as a second-line treatment. A previous study [56] showed that patients on sulfonylureas and insulin had a higher CVD risk than those on GLP1 RA, DPP4i, and SGLT2i, but it found no significant difference for patients on sulfonylureas and insulin. Both studies used composite CVD outcomes as opposed to HF exacerbations specifically. They are particularly important

XSL•FO
RenderX

because frequent exacerbations are the most costly and burdensome aspect of the disease [7,8]; therefore, finding factors that make them less frequent is vital for the health care system.

There were no significant differences in HF outcomes for GLP1 RA and DPP4i, while a trial emulation study reported that patients on GLP1 RA had a lower CVD risk than those on DPP4i [27]. However, their outcome of interest was major adverse cardiovascular events, and not specifically time-to-HF exacerbation. A 2021 systematic review aiming to summarize the effect of GLP-1 RAs compared to DPP-4i showed conflicting results, with one study suggesting the benefit of GLP-1 RAs and another favoring DPP-4i [57], while a 2022 meta-analysis showed a similar risk of HF-related hospitalizations [58].

We included a broader real-world patient population consisting of both patients with and without preexisting HF, while most existing studies targeted a high-risk subpopulation with prior CVDs. Nevertheless, the rates of HF-related hospitalizations in our study were consistent with those found in the literature. HF is associated with an over 50% five-year mortality rate [59], and a 2023 meta-analysis showed a 35.7% (95% CI 27.1%-44.9%) 1-year readmission rate in patients previously admitted for HF [60]. Moreover, a local study conducted in Minnesota showed that among 1077 patients with HF followed over a mean period of 4.7 years, there were 713 HF-related hospitalizations, and 32.3% of the cohort were hospitalized due to HF at least once [61]. Among patients with T2D in the Empagliflozin Cardiovascular Outcome Event (EMPA-REG OUTCOME) randomized clinical trial of empagliflozin, the HF hospitalization rates were lower and equaled 2.7% (empagliflozin) or 4.1% (placebo) over a median follow-up time of 3.1 years [19]. However, these were patients who fulfilled the inclusion criteria and not the real-world population. Studies included in the meta-analysis conducted by Xu et al [58] reported that HF-related hospitalization rates in patients with T2D were between 0.6% and 8.7% per year [11-13] but increased to almost 30% per year if only patients with a previous HF diagnosis were included [11].

The most important predictors in HF outcome models were preexisting HF and the N-terminal pro-B-type natriuretic peptide laboratory tests, a biomarker for HF [62], which are clearly strong predictors of a future HF exacerbation. Propensity score models contained numerous other variables, including many EHR-derived, NLP-derived, and hierarchical medication structure variables. This shows that treatment assignment is a complicated process, which depends on many factors, and obtaining credible results of the final analysis requires state-of-the-art approaches to EHR data. There is no single straightforward T2D treatment regime sequence [18], and in some patients, clinical practitioners choose between insulin and sulfonylureas. This is why a comprehensive comparison between T2D drugs is crucial for clinicians to guide their decisions. In light of our findings, sulfonylureas might be associated with lower risks of HF-related hospitalization compared to insulin, but this needs to be validated by further studies with larger sample sizes and less confounding designs.

Limitations

First, as in most observational studies, unadjusted confounding may exist and compromise the conclusions. Second, the sample sizes were limited by our capacity to manually annotate clinical data and may lead to underpowered analyses. After applying eligibility, the resulting sample sizes could only detect >6.7% risk difference for insulin versus sulfonylureas, and >10.6% risk difference or >12% relative risks for GLP1 RA versus DPP4i. While our subsampling for annotation was completely random and supposedly representative, the eligibility criteria may have introduced selection bias due to imprecise data. We also could not identify patients assigned to combination therapies (eg, metformin and GLP1 RA) based on prescription codes alone. As artificial intelligence language models advance, we envision that artificial intelligence–assisted data abstraction will significantly increase the availability of high-quality data. Furthermore, the data annotated from this study may facilitate the training and validation of these future tools. Third, our intent-to-treat style primary analysis only considered the point decision of treatment with medication initiated after metformin, which did not account for subsequent treatment switching nor investigate sequential treatment strategies. Interpretation of intent-to-treat style analysis is dependent on the subsequent treatment patterns and may not generalize to other institutions or future times [63]. While our sensitivity analysis produced generally lower HF-related hospitalization rates, it may be subject to selection bias if treatment switching was informative for poor HF outcomes. Previous studies [64,65] have proposed the use of composite outcomes involving treatment switching or discontinuation, but the relatively high rate of treatment switching in this study would dominate the composite outcome and deviate substantially from our focus on severe HF outcomes. Similarly, postindex SGLT2i or metformin use may impact the HF outcomes and is difficult to adjust for due to its likely informativeness with HF outcomes. While we observed heterogeneity among treatment groups, the higher postindex SGLT2i use (GLP1 RA and insulin) is associated with a higher estimated HF-related hospitalization (Table S2 in Multimedia Appendix 1; Figure 3). These questions can be revisited once corresponding statistical methods are developed. Finally, the study only included patients from a single institute and lacked the validation of generalizability. Multi-institutional collaboration through research networks may be considered in the future [65].

Conclusions

In conclusion, we created a T2D study cohort from EHRs with annotated clinical data and compared the 5-year HF outcomes of insulin versus sulfonylureas and GLP1 RA versus DPP4i. After adjustment for confounding, we found that sulfonylureas were associated with a reduced risk for HF-related hospitalization compared to insulin, while we could not detect a significant difference for HF outcomes in patients treated with GLP1 RA and DPP4i with the current sample size. Our findings provided additional evidence to guide clinical decisions for managing T2D.

XSL•FO
RenderX

Funding

The study was partially supported by University of Minnesota Data Science Initiative Seed Grant (DSI-MLSG-6875532636).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to laws and regulations on protected health information contained in the datasets but are available from the corresponding author on reasonable request and regulatory approval.

Authors' Contributions

Conceptualization: JH, RZ, RD
Data curation: EJ-S, YC, JL, SZ
Formal analysis: EJ-S, YC, JH
Funding acquisition: JH, RZ
Methodology: JH, EJ-S, YC, JL, SZ
Project administration: JH
Supervision: JH
Visualization: EJ-S, YC, SZ, JH
Writing – original draft: EJ-S, YC, JH
Writing – review & editing: all coauthors

Conflicts of Interest

RD is the editor-in-chief of *JMIR Dermatology*. Other authors have no conflicts to declare.

Multimedia Appendix 1
Secondary and sensitivity analyses results and supplementary figures and tables.
[DOCX File, 752 KB - diabetes_v11i1e85083_app1.docx]

References

1. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
2. National diabetes statistics report. : Centers for Disease Control and Prevention URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2025-07-01]
3. Cho NH, Shaw JE, Karuranga S, et al. IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
4. Booth GL, Kapral MK, Fung K, Tu JV. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. Lancet 2006 Jul 1;368(9529):29-36. [doi: 10.1016/S0140-6736(06)68967-8] [Medline: 16815377]
5. Dal Canto E, Ceriello A, Rydén L, et al. Diabetes as a cardiovascular risk factor: an overview of global trends of macro and micro vascular complications. Eur J Prev Cardiol 2019 Dec;26(2_suppl):25-32. [doi: 10.1177/2047487319878371] [Medline: 31722562]
6. Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical update: cardiovascular disease in diabetes mellitus: atherosclerotic cardiovascular disease and heart failure in type 2 diabetes mellitus - mechanisms, management, and clinical considerations. Circulation 2016 Jun 14;133(24):2459-2502. [doi: 10.1161/CIRCULATIONAHA.116.022194] [Medline: 27297342]
7. Jerkins T, McGill JB, Bell DS. Heart failure and diabetes: clinical significance and epidemiology of this two-way association. Diabetes Obes Metab 2023 Jul;25 Suppl 3(S3):3-14. [doi: 10.1111/dom.15062] [Medline: 36949650]
8. Osenenko KM, Kuti E, Deighton AM, Pimple P, Szabo SM. Burden of hospitalization for heart failure in the United States: a systematic literature review. J Manag Care Spec Pharm 2022 Feb;28(2):157-167. [doi: 10.18553/jmcp.2022.28.2.157] [Medline: 35098748]
9. Belfiore A, Stranieri R, Novielli ME, Portincasa P. Reducing the hospitalization epidemic of chronic heart failure by disease management programs. Intern Emerg Med 2024 Jan;19(1):221-231. [doi: 10.1007/s11739-023-03458-7] [Medline: 38151590]
10. Hospitalization. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/hus/topics/hospitalization.htm [accessed 2025-07-01]
11. Dawwas GK, Smith SM, Park H. Risk of heart failure hospitalization among users of dipeptidyl peptidase-4 inhibitors compared to glucagon-like peptide-1 receptor agonists. Cardiovasc Diabetol 2018 Jul 17;17(1):102. [doi: 10.1186/s12933-018-0746-4] [Medline: 30016946]

XSL•FO
RenderX

12. Svanström H, Ueda P, Melbye M, et al. Use of liraglutide and risk of major cardiovascular events: a register-based cohort study in Denmark and Sweden. Lancet Diabetes Endocrinol 2019 Feb;7(2):106-114. [doi: 10.1016/S2213-8587(18)30320-6] [Medline: 30527909]
13. Longato E, Di Camillo B, Sparacino G, Gubian L, Avogaro A, Fadini GP. Cardiovascular outcomes of type 2 diabetic patients treated with SGLT-2 inhibitors versus GLP-1 receptor agonists in real-life. BMJ Open Diabetes Res Care 2020 Jun;8(1):e001451. [doi: 10.1136/bmjdrc-2020-001451] [Medline: 32591373]
14. Nair R, Lak H, Hasan S, Gunasekaran D, Babar A, Gopalakrishna KV. Reducing all-cause 30-day hospital readmissions for patients presenting with acute heart failure exacerbations: a quality improvement initiative. Cureus 2020 Mar 25;12(3):e7420. [doi: 10.7759/cureus.7420] [Medline: 32351805]
15. American Diabetes Association Professional Practice Committee for Diabetes. 9. Pharmacologic approaches to glycemic treatment: standards of care in diabetes-2026. Diabetes Care 2026 Jan 1;49(Supplement_1):S183-S215. [doi: 10.2337/dc26-S009] [Medline: 41358900]
16. Jasleen B, Vishal GK, Sameera M, et al. Sodium-glucose cotransporter 2 (SGLT2) inhibitors: benefits versus risk. Cureus 2023 Jan 18;15(1):e33939. [doi: 10.7759/cureus.33939] [Medline: 36819350]
17. Ussher JR, Drucker DJ. Glucagon-like peptide 1 receptor agonists: cardiovascular benefits and mechanisms of action. Nat Rev Cardiol 2023 Jul;20(7):463-474. [doi: 10.1038/s41569-023-00849-3] [Medline: 36977782]
18. Samson SL, Vellanki P, Blonde L, et al. American Association of Clinical Endocrinology consensus statement: comprehensive type 2 diabetes management algorithm - 2023 update. Endocr Pract 2023 May;29(5):305-340. [doi: 10.1016/j.eprac.2023.02.001] [Medline: 37150579]
19. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015 Nov 26;373(22):2117-2128. [doi: 10.1056/NEJMoa1504720] [Medline: 26378978]
20. Fitchett D, Zinman B, Wanner C, et al. Heart failure outcomes with empagliflozin in patients with type 2 diabetes at high cardiovascular risk: results of the EMPA-REG OUTCOME® trial. Eur Heart J 2016 May 14;37(19):1526-1534. [doi: 10.1093/eurheartj/ehv728] [Medline: 26819227]
21. Rådholm K, Figtree G, Perkovic V, et al. Canagliflozin and heart failure in type 2 diabetes mellitus: results from the CANVAS Program. Circulation 2018 Jul 31;138(5):458-468. [doi: 10.1161/CIRCULATIONAHA.118.034222] [Medline: 29526832]
22. Kosiborod MN, Abildstrøm SZ, Borlaug BA, et al. Semaglutide in patients with heart failure with preserved ejection fraction and obesity. N Engl J Med 2023 Sep 21;389(12):1069-1084. [doi: 10.1056/NEJMoa2306963] [Medline: 37622681]
23. Scirica BM, Braunwald E, Raz I, et al. Heart failure, saxagliptin, and diabetes mellitus: observations from the SAVOR-TIMI 53 randomized trial. Circulation 2014 Oct 28;130(18):1579-1588. [doi: 10.1161/CIRCULATIONAHA.114.010389] [Medline: 25189213]
24. Khan MS, Fonarow GC, McGuire DK, et al. Glucagon-like peptide 1 receptor agonists and heart failure: the need for further evidence generation and practice guidelines optimization. Circulation 2020 Sep 22;142(12):1205-1218. [doi: 10.1161/CIRCULATIONAHA.120.045888] [Medline: 32955939]
25. Franklin JM, Liaw KL, Iyasu S, Critchlow CW, Dreyer NA. Real‐world evidence to support regulatory decision making: new or expanded medical product indications. Pharmacoepidemiol Drug Saf 2021 Jun;30(6):685-693. [doi: 10.1002/pds.5222] [Medline: 33675248]
26. Hou J, Zhao R, Gronsbell J, et al. Generate analysis-ready data for real-world evidence: tutorial for harnessing electronic health records with advanced informatic technologies. J Med Internet Res 2023 May 25;25:e45662. [doi: 10.2196/45662] [Medline: 37227772]
27. Xie Y, Bowe B, Xian H, Loux T, McGill JB, Al-Aly Z. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of major adverse cardiovascular events: emulation of a randomised target trial using electronic health records. Lancet Diabetes Endocrinol 2023 Sep;11(9):644-656. [doi: 10.1016/S2213-8587(23)00171-7] [Medline: 37499675]
28. Wang X, Plantinga AM, Xiong X, et al. Comparing insulin against glucagon-like peptide-1 receptor agonists, dipeptidyl peptidase-4 inhibitors, and sodium-glucose cotransporter 2 inhibitors on 5-year incident heart failure risk for patients with type 2 diabetes mellitus: real-world evidence study using insurance claims. JMIR Diabetes 2024 Oct 22;9:e58137. [doi: 10.2196/58137] [Medline: 39436276]
29. Giorgino F, Leonardini A, Laviola L. Cardiovascular disease and glycemic control in type 2 diabetes: now that the dust is settling from large clinical trials. Ann N Y Acad Sci 2013 Apr;1281(1):36-50. [doi: 10.1111/nyas.12044] [Medline: 23387439]
30. Savitz ST, Savitz LA, Fleming NS, Shah ND, Go AS. How much can we trust electronic health record data? Healthc (Amst) 2020 Sep;8(3):100444. [doi: 10.1016/j.hjdsi.2020.100444] [Medline: 32919583]
31. Hong C, Rush E, Liu M, et al. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. NPJ Digit Med 2021 Oct 27;4(1):151. [doi: 10.1038/s41746-021-00519-z] [Medline: 34707226]
32. Xiong X, Sweet SM, Liu M, et al. Knowledge-driven online multimodal automated phenotyping system. medRxiv. 2025 Mar 5 p. 2023.09.29.23296239. [doi: 10.1101/2023.09.29.23296239] [Medline: 37873131]

XSL•FO
RenderX

33. Horsky J, Drucker EA, Ramelson HZ. Accuracy and completeness of clinical coding using ICD-10 for ambulatory visits. AMIA Annu Symp Proc 2018;2017:912-920. [Medline: 29854158]
34. Weiskopf NG, Cohen AM, Hannan J, Jarmon T, Dorr DA. Towards augmenting structured EHR data: a comparison of manual chart review and patient self-report. AMIA Annu Symp Proc 2020 Mar 4;2019:903-912. [Medline: 32308887]
35. Zhang Y, Cai T, Yu S, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat Protoc 2019 Dec;14(12):3426-3444. [doi: 10.1038/s41596-019-0227-6] [Medline: 31748751]
36. Huang S, Cai T, Weber BN, et al. Association between inflammation, incident heart failure, and heart failure subtypes in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2023 May;75(5):1036-1045. [doi: 10.1002/acr.24804] [Medline: 34623035]
37. Siems A, Banks R, Holubkov R, et al. Structured chart review: assessment of a structured chart review methodology. Hosp Pediatr 2020 Jan;10(1):61-69. [doi: 10.1542/hpeds.2019-0225] [Medline: 31879317]
38. Okura T, Fujioka Y, Nakamura R, et al. Dipeptidyl peptidase 4 inhibitor improves insulin resistance in Japanese patients with type 2 diabetes: a single-arm study, a brief report. Diabetol Metab Syndr 2022 Jun 7;14(1):78. [doi: 10.1186/s13098-022-00850-9] [Medline: 35672759]
39. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics 2005 Dec;61(4):962-973. [doi: 10.1111/j.1541-0420.2005.00377.x] [Medline: 16401269]
40. Hou J, Bradic J, Xu R. Treatment effect estimation under additive hazards models with high-dimensional confounding. J Am Stat Assoc 2023 Jan 2;118(541):327-342. [doi: 10.1080/01621459.2021.1930546]
41. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010 May 1;26(9):1205-1210. [doi: 10.1093/bioinformatics/btq126] [Medline: 20335276]
42. Liao KP, Sun J, Cai TA, et al. High-throughput multimodal automated phenotyping (MAP) with application to PheWAS. J Am Med Inform Assoc 2019 Nov 1;26(11):1255-1262. [doi: 10.1093/jamia/ocz066] [Medline: 31613361]
43. Hou J, Guo Z, Cai T. Surrogate assisted semi-supervised inference for high dimensional risk prediction. J Mach Learn Res 2023;24:265. [Medline: 38500567]
44. Yu S, Cai T, Cai T. NILE: fast natural language processing for electronic health records. arXiv. Preprint posted online on Nov 23, 2013. [doi: 10.48550/arXiv.1311.6063]
45. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270. [doi: 10.1093/nar/gkh061] [Medline: 14681409]
46. Nessler J. Dapagliflozin in the treatment of patients with heart failure with reduced left ventricular ejection fraction – a practical approach. Postepy Kardiol Interwencyjnej 2021 Jun;17(2):135-140. [doi: 10.5114/aic.2021.107491] [Medline: 34400915]
47. Teo YH, Teo YN, Syn NL, et al. Effects of sodium/glucose cotransporter 2 (SGLT2) inhibitors on cardiovascular and metabolic outcomes in patients without diabetes mellitus: a systematic review and meta-analysis of randomized-controlled trials. J Am Heart Assoc 2021 Feb;10(5):e019463. [doi: 10.1161/JAHA.120.019463] [Medline: 33625242]
48. Giles TD, Miller AB, Elkayam U, Bhattacharya M, Perez A. Pioglitazone and heart failure: results from a controlled study in patients with type 2 diabetes mellitus and systolic dysfunction. J Card Fail 2008 Aug;14(6):445-452. [doi: 10.1016/j.cardfail.2008.02.007] [Medline: 18672190]
49. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc 2011;18(4):441-448. [doi: 10.1136/amiajnl-2011-000116] [Medline: 21515544]
50. Hou J, Mukherjee R, Cai T. Efficient and robust semi-supervised estimation of average treatment effect with partially annotated treatment and response. J Mach Learn Res 2025;26:40. [Medline: 41341524]
51. Huff SM, Rocha RA, McDonald CJ, et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc 1998;5(3):276-292. [doi: 10.1136/jamia.1998.0050276] [Medline: 9609498]
52. Clinical Classifications Software for Services and Procedures. US Agency for Healthcare Research and Quality. 2014. URL: https://hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp [accessed 2025-07-01]
53. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994 Sep;89(427):846-866. [doi: 10.1080/01621459.1994.10476818]
54. Zou H. The adaptive lasso and its oracle properties. J Am Stat Assoc 2006;101(476):1418-1429. [doi: 10.1198/016214506000000735]
55. Gebrie D, Manyazewal T, Ejigu DA, Makonnen E. Metformin-insulin versus metformin-sulfonylurea combination therapies in type 2 diabetes: a comparative study of glycemic control and risk of cardiovascular diseases in Addis Ababa, Ethiopia. Diabetes Metab Syndr Obes 2021 Jul 24;14:3345-3359. [doi: 10.2147/DMSO.S312997] [Medline: 34335036]
56. O'Brien MJ, Karam SL, Wallia A, et al. Association of second-line antidiabetic medications with cardiovascular events among insured adults with type 2 diabetes. JAMA Netw Open 2018 Dec 7;1(8):e186125. [doi: 10.1001/jamanetworkopen.2018.6125] [Medline: 30646315]

XSL•FO
RenderX

57. Alkhezi OS, Alsuhaibani HA, Alhadyab AA, et al. Heart failure outcomes and glucagon-like peptide-1 receptor agonists: a systematic review of observational studies. Prim Care Diabetes 2021 Oct;15(5):761-771. [doi: 10.1016/j.pcd.2021.04.005] [Medline: 33926837]
58. Xu L, Zheng XQ, Liao XX. Cardiovascular effectiveness of glucagon-like peptide 1 receptor agonists versus dipeptidyl peptidase-4 inhibitors in type 2 diabetes: a meta-analysis based on propensity score-matched studies. Prim Care Diabetes 2022 Feb;16(1):207-210. [doi: 10.1016/j.pcd.2021.11.006] [Medline: 34953749]
59. Jones NR, Roalfe AK, Adoki I, Hobbs FD, Taylor CJ. Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis. Eur J Heart Fail 2019 Nov;21(11):1306-1325. [doi: 10.1002/ejhf.1594] [Medline: 31523902]
60. Foroutan F, Rayner DG, Ross HJ, et al. Global comparison of readmission rates for patients with heart failure. J Am Coll Cardiol 2023 Aug 1;82(5):430-444. [doi: 10.1016/j.jacc.2023.05.040] [Medline: 37495280]
61. Dunlay SM, Redfield MM, Weston SA, et al. Hospitalizations after heart failure diagnosis a community perspective. J Am Coll Cardiol 2009 Oct 27;54(18):1695-1702. [doi: 10.1016/j.jacc.2009.08.019] [Medline: 19850209]
62. Mueller C, McDonald K, de Boer RA, et al. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. Eur J Heart Fail 2019 Jun;21(6):715-731. [doi: 10.1002/ejhf.1494] [Medline: 31222929]
63. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017 Oct 5;377(14):1391-1398. [doi: 10.1056/NEJMsm1605385] [Medline: 28976864]
64. Heerspink HJ, Karasik A, Thuresson M, et al. Kidney outcomes associated with use of SGLT2 inhibitors in REAL-world clinical practice (CVD-REAL 3): a multinational observational cohort study. Lancet Diabetes Endocrinol 2020 Jan;8(1):27-35. [doi: 10.1016/S2213-8587(19)30384-5] [Medline: 31862149]
65. Khera R, Schuemie MJ, Lu Y, et al. Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies. BMJ Open 2022 Jun 9;12(6):e057977. [doi: 10.1136/bmjopen-2021-057977] [Medline: 35680274]

Abbreviations

CVD: cardiovascular disease
DPP4i: dipeptidyl peptidase-4 inhibitors
EHR: electronic health record
GLP1 RA: glucagon-like peptide-1 receptor agonists
HF: heart failure
MAP: multimodal automated phenotyping
NLP: natural language processing
SGLT2i: sodium-glucose cotransporter 2 inhibitors
T2D: type 2 diabetes mellitus

Edited by S Li; submitted 04.Oct.2025; peer-reviewed by C Xiang, YS Liu; revised version received 18.Feb.2026; accepted 20.Feb.2026; published 15.Apr.2026.

Please cite as:
Jodlowska-Siewert E, Chen Y, Zhang S, Li J, Dellavalle R, Zhang R, Hou J
Antidiabetic Drug Associations With Heart Failure Outcomes: Real-World Evidence Study Using Electronic Health Records
JMIR Diabetes 2026;11:e85083
URL: https://diabetes.jmir.org/2026/1/e85083
doi:10.2196/85083

© Elzbieta Jodlowska-Siewert, Yunhui Chen, Sinian Zhang, Jia Li, Robert Dellavalle, Rui Zhang, Jue Hou. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 15.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cardiorespiratory Markers of Type 2 Diabetes: Machine Learning–Based Analysis

Flavia Maria G S A Oliveira[1], PhD; Sandro Muniz Cavalcanti[1], BSc; Michael C K Khoo[2], PhD

[1]Department of Electrical Engineering, School of Technology, University of Brasilia, Campus Universitário Darcy Ribeiro, Asa Norte, Brasilia-DF, Brazil

[2]Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Flavia Maria G S A Oliveira, PhD
Department of Electrical Engineering, School of Technology, University of Brasilia, Campus Universitário Darcy Ribeiro, Asa Norte, Brasilia-DF, Brazil

Abstract

Background: The global prevalence of type 2 diabetes mellitus (T2DM) poses significant challenges due to its association with increased cardiovascular risk and complications like cardiovascular autonomic neuropathy. Measures derived from heart rate variability (HRV) and cardiorespiratory interactions quantified through frequency response function (FRF) and impulse response (IR) metrics reflect different aspects of autonomic regulation and may provide complementary physiological information relevant to diabetes-related autonomic alterations.

Objective: The study aimed to investigate whether these metrics, individually or in combination, provide useful physiological features for distinguishing individuals with and without T2DM using machine learning classifiers.

Methods: Electrocardiogram and respiratory signals from 2 PhysioNet datasets were used to derive 3 domains of autonomic and cardiorespiratory features: (1) spectral HRV indices reflecting overall variability; (2) FRF metrics characterizing frequency-specific respiratory-cardiac transfer properties; and (3) causal IR metrics capturing time-domain responsiveness to respiratory inputs. ML classifiers—logistic regression, support vector machine (SVM) with linear kernel, and SVM with radial basis function (SVM RBF) kernel—assessed the predictive value of individual and combined feature sets under NearMiss-1 (NM) undersampling and Synthetic Minority Oversampling Technique oversampling. This systems-based framework may capture subtle differences in respiratory-cardiac regulation associated with T2DM more effectively than HRV alone by reflecting integrated cardiorespiratory coupling.

Results: Across classifiers and balancing strategies, IR features frequently produced comparatively strong standalone performance, suggesting that causal, time-domain cardiorespiratory dynamics capture informative physiological differences between groups. With logistic regression and NM, IR features achieved mean accuracy of 0.770 (SD 0.179), precision of 0.783 (SD 0.217), recall of 0.900 (SD 0.224), and F_1-score of 0.798 (SD 0.140). While HRV metrics were the least informative standalone feature set, the combined HRV+FRF feature set under NM yielded the highest observed performance, with accuracy of 0.830 (SD 0.172), precision of 0.800 (SD 0.183), recall of 0.933 (SD 0.149), and F_1-score of 0.853 (SD 0.145; SVM RBF). Under Synthetic Minority Oversampling Technique, HRV+IR showed the strongest observed combined performance, yielding accuracy of 0.700 (SD 0.128), precision of 0.783 (SD 0.217), recall of 0.683 (SD 0.207), and F_1-score of 0.691 (SD 0.097) with SVM RBF, surpassing standalone IR in most metrics, though IR alone retained superior recall (0.950, SD 0.112) and F_1-score (0.708, SD 0.038). These results reflect that performance depends on both feature domain and sampling strategy and that combining features capturing complementary physiological aspects of autonomic regulation may enhance discriminative ability.

Conclusions: HRV, FRF, and IR metrics each reflect distinct dimensions of autonomic and cardiorespiratory regulation. Systems-based approaches incorporating frequency-domain and causal dynamic features may offer richer characterization of diabetes-related regulatory differences than HRV alone. Although preliminary and limited by sample size, these findings highlight promising physiological feature domains and sampling strategies for future investigation. Larger datasets with well-defined autonomic phenotyping are needed to evaluate generalizability and determine clinical relevance.

(JMIR Diabetes 2026;11:e82084) doi:10.2196/82084

KEYWORDS

heart rate variability; frequency response function; impulse response; type 2 diabetes; diabetic autonomic neuropathy; machine learning; cardiorespiratory coupling

XSL•FO
RenderX

Introduction

The International Diabetes Federation [1] reports that global prevalence of diabetes in adults aged 20 - 79 years rose from 151 million (about 4.6% of the global population) in 2000, to 589 million (11.1%) in 2024, with projections reaching 853 million (13.0%) by 2050. Over 90% of these cases are type 2 diabetes mellitus (T2DM). Diabetes is associated with a twofold increase in the risk of vascular diseases, including coronary heart disease and stroke, independent of common risk factors such as obesity and hypertension [2].

Cardiovascular autonomic neuropathy (CAN), affecting from 12% to 73% of patients with T2DM and linked to a 3.45-fold mortality risk [3], is a highly prevalent but frequently overlooked microvascular complication, also found in prediabetic individuals with metabolic syndrome [4]. Sörensen et al [5] demonstrated that generalized microvascular dysfunction is already present in prediabetes and becomes more pronounced in established T2DM. These findings suggest that microvascular impairment precedes and may contribute to later cardiovascular complications in T2DM, supporting its role as a potential early target for intervention.

Reduced heart rate variability (HRV) is an early sign of CAN in individuals with diabetes and prediabetes [6]. HRV has been used in many studies to quantify cardiovascular autonomic function in T2DM, demonstrating that diabetes significantly affects both the sympathetic and parasympathetic branches of the autonomic nervous system [7]. Lower HRV in diabetic subjects, compared to controls, indicates an impaired ability to adapt to physiological stressors, such as physical exercise and orthostatic stress. Using HRV to measure cardiac autonomic function, Wang et al [8] demonstrated that autonomic dysfunction precedes the development of T2DM, particularly in younger individuals, even after adjusting for cardiovascular risk factors. Given that HRV metrics can capture changes in autonomic control associated with diabetes progression [9], they may provide insight into early physiological alterations in individuals at risk for T2DM. Some studies have advocated the use of standardized cardiovascular autonomic reflex tests (CARTs) rather than measurements of spontaneous HRV, but a major limitation of CARTs is that they require varying levels of subject cooperation, and differing levels of attention or anxiety may lead to intersubject and intrasubject variability in test results [10].

While HRV analysis offers important insights into cardiovascular autonomic regulation, it is well known that respiratory patterns influence HRV (eg, respiratory sinus arrhythmia), potentially confounding cardiovascular autonomic assessments. Different approaches have been proposed in the literature to address this issue. One approach is to estimate the frequency response function (FRF), or spectral transfer function, between changes in respiration (as input) to variations in R-to-R interval (RRI) (as output) [11-13]. For instance, Khoo et al [11] demonstrated that the average transfer gain between respiration and RRI, an FRF-derived metric representing vagal control that specifically accounts for respiratory influences, outperformed traditional HRV by explicitly accounting for respiratory contributions. In the context of T2DM, such FRF-based descriptors may offer complementary insights into cardiorespiratory interactions that differ between individuals with and without diabetes, highlighting physiological dimensions not captured by HRV alone.

Since the cardiovascular and respiratory systems are inherently interconnected in a closed-loop framework, a limitation of the FRF approach is its assumption of a 1-directional, feedforward influence from input to output, with no reciprocal feedback. However, these systems interact bidirectionally, with cardiovascular and respiratory signals influencing each other. This simplification may complicate the interpretation of results. To address the inherent noncausal nature of the FRF, an alternative approach is to estimate the time-domain impulse response (IR) between measured input and output within a mathematical model of the underlying dynamics. This representation allows the output to depend explicitly on present and past values of the input, but not future values, as well as the inclusion of delays into the model. This effectively "opens the loop," helping separate feedforward influences from feedback interactions [14,15].

In this study, we sought to investigate how these different, complementary measures of autonomic regulation—spectral measures of HRV, noncausal systems-based FRF metrics, and causal IR-derived indices—differ between subjects with and without T2DM. We then examined whether these physiologically grounded descriptors provide discriminative value for distinguishing T2DM from controls in machine learning (ML) models.

To the best of our knowledge, this is the first study to evaluate HRV, FRF, and IR metrics collectively in a ML framework to examine physiologically grounded differences between individuals with and without T2DM. By systematically evaluating individual and combined feature domains, this study provides exploratory evidence on the potential value of multivariate, systems-based physiological descriptors in distinguishing diabetes-related regulatory patterns.

Methods

Database

This study used 2 publicly available PhysioNet repositories curated by the same research group: the cerebral vasoregulation in diabetes dataset [16] and the cerebral perfusion and cognitive decline in type 2 diabetes dataset [17]. Both datasets were collected at the Syncope and Falls in the Elderly Laboratory at Beth Israel Deaconess Medical Center (BIDMC), Harvard Medical School, Boston, MA, by the research group of Dr. Vera Novak, under comparable experimental conditions and using the same core infrastructure, ensuring methodological compatibility. Among other measures, the studies analyzed electrocardiogram (ECG) and respiration signals recorded during a standardized "sit-to-stand" test in subjects with T2DM and age-matched controls, aged 50 - 85 years.

The demographics and clinical characteristics, including hemoglobin A1c, of the groups that participated in the sit-to-stand test in both datasets are summarized in Table 1,

XSL•FO
RenderX

reflecting mean and SD of available data, as some participants lacked demographic (n=2) or clinical (n=8) information.

Table . Main characteristics of the groups.

Variable	Groups, mean (SD)		*P* value
	Control (n=21)	T2DM[a] (n=49)	
Age (y)	64.6 (8.1)	63.7 (8.1)	.61
Mass (kg)	68.6 (11.2)	83.2 (15)	<.001[b]
BMI	24.2 (2.6)	29.0 (5.1)	<.001[b]
HbA_{1c}[c] (%)	5.4 (0.4)	7.2 (1.4)	<.001[b]
RRI[d] (ms)	877.4 (144.9)	825.8 (137.8)	<.001[b]

[a]T2DM: type 2 diabetes mellitus.

[b]$P \leq .05$ (data are mean [SD]).

[c]HbA_{1c}: hemoglobin A_{1c}.

[d]RRI: R-to-R interval.

Ethical Considerations

This study involved secondary analysis of two publicly available, deidentified datasets hosted on PhysioNet: the Cerebral Vasoregulation in Diabetes dataset and the Cerebral Perfusion and Cognitive Decline in Type 2 Diabetes dataset. The original data collections were conducted at BIDMC under institutional review board (IRB) approval (IRB 2003P000013 and IRB 2005P000338, respectively). All participants provided written informed consent prior to enrollment in the original studies.

In the original protocols, participants were admitted to the Clinical Research Center at BIDMC, where all study procedures were conducted under medical supervision. Privacy and confidentiality were maintained under IRB-approved procedures, and data were handled in accordance with institutional and federal regulations governing human subjects research. The datasets made available on PhysioNet were fully deidentified prior to public release, and no personally identifiable information is included.

This study used only publicly available, deidentified data and did not involve direct contact with participants. Therefore, no additional ethics approval or informed consent was required for this secondary analysis.

The original study documentation does not specify whether participants received financial compensation for participation.

All procedures adhered to the ethical standards of the responsible IRBs and to the principles outlined in the Declaration of Helsinki.

Data Preprocessing

For accurate interpretation of spectral measures of HRV, the data should be essentially stationary. As per the European Society of Cardiology/North American Society of Pacing and Electrophysiology Task Force on HRV [18], recordings should be short enough to meet stationarity requirements for frequency-domain analyses, yet long enough to capture at least 10 cycles of the cut-off frequency for the low-frequency (LF) HRV component, typically set at 0.04 Hz (corresponding to a periodicity of 25 s). Taking these considerations into account, we selected 4-minute (240 s) segments in the sitting position for analysis. Recordings with significant signal loss due to equipment recalibration, brief unexplained signal flattening (repeatedly found in the respiratory signals), or excessive ectopic beats were excluded from analysis. In the recordings that were available for this study, only about half of the participants had data measured in both sitting and standing postures. To maximize sample size for our analyses, we chose to use data from the sitting posture only.

The original pooled dataset included 21 control subjects and 49 individuals with T2DM. However, because the IR feature set requires each subject to have both valid ECG and respiration signals, we restricted analysis to subjects who had valid recordings of both signals. To ensure directly comparable evaluations of all feature sets (HRV, FRF, and IR), we elected to use the same subset of participants across every machine-learning analysis. After applying all preprocessing steps, the final consistent dataset comprised 18 T2DM subjects and 11 controls. These fixed sample sizes were used for all classifiers and all feature-set comparisons to avoid bias introduced by varying subject availability across methods.

Data processing was performed using the Cardiorespiratory System Identification Lab [19], a freely available MATLAB-based software tool for evaluating autonomic nervous system function through HRV and cardiorespiratory system analysis. Key steps included detecting R-waves in the ECG using a Pan-Tompkins-based algorithm [20] to obtain the RRI time series and converting the airflow data (in mL/s) into instantaneous lung volume (ILV) in mL. To prepare the data for spectral and IR analyses, the RRI and ILV signals were resampled at 4 Hz after detrending to eliminate very LF oscillations [21].

Spectral Analysis of HRV

Spectral analysis of the resampled HRV signal was conducted using power spectral density estimation via the Welch method with a Hann window to minimize spectral leakage. We

XSL•FO
RenderX

calculated the LF (0.04 to 0.15 Hz) and high-frequency (HF) (0.15 to 0.4 Hz) [18] components of the RRI time series, as well as the LF by HF ratio. These spectral indices provide insights into cardiac autonomic modulation, with HRVHF often linked to vagal activity and HRVLF reflecting a mix of sympathetic and parasympathetic inputs [18,22-24], with the ratio HRVLF/HF(HRVLF/HRVHF) commonly interpreted as a measure of "sympathovagal balance," although this view has been challenged [25-27].

System-Based Analyses in the Frequency and Time Domains

To incorporate the influence of respiratory-heart rate coupling, we used the FRF and IR analyses. FRF estimates how an output response (eg, RRI) is modulated by an input (eg, respiration), providing a frequency-based perspective on autonomic regulation [28]. Specifically, the FRF gain quantifies the efficiency of coupling between respiratory inputs and cardiac responses, with higher gains indicating stronger modulation of heart rate by respiratory oscillations [28]. In this study, we calculated FRF gain values for LF and HF bands (|FRF|LF and |FRF|HF, respectively) to quantify respiratory influences on RRI. By analyzing frequency-specific dynamics, the FRF highlights how respiratory-cardiac coupling (RCC) varies across physiologically relevant frequency bands. However, FRF is inherently limited in its ability to assess causal interactions, as it does not disentangle feedforward from feedback mechanisms or establish directionality [29].

To address these limitations, IR analysis was used to provide a time-domain, causal perspective on the system's dynamics. By modeling the system's response to an impulse input in respiration, IR analysis allows for the characterization of the system's adaptability, assessing how effectively and over what time frame the cardiovascular system can adjust to respiratory inputs or other perturbations.

To quantify the ILV-to-RRI IR, we calculated several key descriptors: IR magnitude, which reflects the strength of the immediate respiratory influence on cardiac output; dynamic gain (DG, using total, LF, and HF components), which represents the average magnitude of the system's influence across different frequency bands; and characteristic time (t_{char}), which captures the time it takes for the response to occur or subside, providing insights into delayed or sustained effects.

These descriptors facilitate statistical comparisons between groups and capture essential regulatory characteristics of the RCC mechanism [15]. Together, these metrics help reveal the system's ability to maintain stability and recover from changes, offering critical insights into the system's flexibility and robustness in both health and disease [15,30]. Further details on the FRF and IR methodologies are available in the online supplementary material (Multimedia Appendix 1).

Analysis Procedures

Machine Learning Classifiers

To distinguish T2DM subjects from controls, we used three ML classifiers: logistic regression (LR), support vector machines (SVMs) with linear kernels (SVM linear), and SVM with radial basis function (SVM RBF kernels). We trained and tested these classifiers using various feature sets—as described below—which capture distinct but complementary aspects of autonomic nervous system function.

Feature Sets and Groupings

Although HRV, FRF, and IR features all originate from the cardiac timing signal (RRI), they differ in the extent to which they incorporate respiratory information and in the physiological mechanisms they reflect. Spectral HRV indices (LF, HF, and LF/HF) quantify the distribution of oscillatory RRI variability but remain univariate summary descriptors of oscillatory patterns influenced by multiple regulatory pathways. FRF metrics quantify the frequency-specific transfer characteristics of respiratory—cardiac interactions, capturing the gain and phase relationships that reflect how respiratory oscillations shape cardiac timing, but without modeling causal direction. IR metrics extend this systems-based perspective by characterizing the causal, time-domain responsiveness of RRI to respiratory perturbations, thereby providing information about dynamic adaptability and directional regulation that is not accessible from HRV or FRF measures alone. Given these complementary perspectives—overall variability patterns (HRV), frequency-dependent transfer behavior (FRF), and causal dynamics responsiveness (IR)—we additionally evaluated whether combining HRV with FRF or IR features provided complementary discriminative information beyond any single feature domain.

Thus, initially, each classification model was trained on one of the following individual feature sets: (a) HRV metrics, (b) FRF metrics, or (c) IR metrics. To assess whether combining feature sets could enhance classification performance, we created additional feature groupings: (d) HRV+FRF metrics and (e) HRV+IR metrics.

Covariates, including BMI, were not incorporated into the ML models. The present study was designed to examine the physiological content and discriminative behavior of HRV, FRF, and IR features, rather than to develop covariate-adjusted predictive models. This approach reflects the mechanistic focus of this study and avoids the added model complexity and potential instability that covariate adjustment would introduce given the modest sample size.

Handling Class Imbalance

To address the T2DM majority (accounting for about 2/3 of subjects), we applied NearMiss-1 (NM) undersampling [31] and Synthetic Minority Oversampling Technique (SMOTE) [32], comparing performance against the unbalanced dataset. This approach allowed us to evaluate classifier performance under different class balance conditions and to compare the relative effectiveness of undersampling and oversampling in improving predictive accuracy. A balanced training dataset in ML is commonly used to reduce bias toward the majority class.

In this exploratory study, NM and SMOTE were applied to the full usable dataset prior to generating the 5-fold stratified cross-validation partitions, rather than separately within each training fold. This design allowed for direct comparison of balancing strategies under fixed class distributions, but it also

XSL•FO
RenderX

entails that, particularly for SMOTE, synthetic samples were generated using neighborhood information from the entire dataset. As a result, some synthetic samples may appear in both training and test folds. The resulting performance estimates should therefore be viewed within the exploratory, hypothesis-generating scope of the study.

With the usable dataset consisting of 18 T2DM and 11 control subjects, NM undersampling reduced the majority class to match the minority class, yielding 11 T2DM and 11 control subjects (N=22 total). In contrast, SMOTE oversampling synthesized 7 new control samples, producing a balanced dataset of 18 T2DM and 18 controls (N=36 total). All classification models under each balancing strategy were trained using these corresponding sample sizes.

Data Preprocessing and Feature Standardization

Given that features in our dataset span different units, all features were standardized using *z*-score normalization (mean 0, SD 1). This preprocessing step ensures that features with larger values do not dominate the model, which is particularly relevant for distance-based methods like SVM [33]. Standardizing features also improves interpretability in LR, as larger coefficients indicate higher feature importance in classification [34,35].

As with the resampling procedures, *z*-score normalization was applied once, before cross-validation, to maintain consistent feature scaling across all model comparisons. This approach trades off strict fold-wise isolation for stability in a small dataset.

Cross-Validation and Feature Correlation

Prior to model training, we performed correlation analysis within each feature set to identify and exclude features with correlations above 0.8. This minimized redundancy and mitigated multicollinearity, promoting stable classification model estimates and clearer interpretation of feature contributions [36,37]. Applying this filtering prior to cross-validation ensured consistent feature definitions across classifiers and balancing strategies, which would not have been feasible with fold-wise filtering given the modest sample size. Models were then evaluated using 5-fold stratified cross-validation to obtain a more robust estimate of classification performance and reduce the risk of overfitting [38].

Performance Metrics and Evaluation

Each classifier's performance was assessed via accuracy, precision, sensitivity, specificity, F_1-score, and area under the receiver operating characteristic curve (AUC-ROC), averaged across 5-fold cross-validation. Each metric offers unique insights into different aspects of classifier performance:

- Accuracy measures the overall correctness of the classifier by calculating the proportion of correctly classified instances (both true positives and true negatives) among all instances; however, it may be misleading in imbalanced datasets, where the majority class dominates the metric.
- Precision evaluates the proportion of true positive predictions among all positive predictions (true positives + false positives), highlighting the classifier's ability to avoid false positives.
- Sensitivity (recall) reflects the model's capacity to identify true positives among all actual positives (true positives + false negatives), a key metric in clinical contexts where missing true positive cases (false negatives) can be costly.
- Specificity assesses the model's performance in correctly identifying true negatives among all actual negatives (true negatives + false positives), important for determining how well the classifier avoids false positives.
- F_1-score is the harmonic mean of precision and sensitivity, providing a single balanced metric that is useful when classes are imbalanced.
- AUC-ROC summarizes the trade-off between sensitivity (true positive rate) and 1−specificity (false positive rate) across different decision thresholds, indicating the classifier's ability to distinguish between classes, with a higher AUC-ROC value reflecting better overall performance.

Considering multiple performance metrics provides a comprehensive assessment of each classifier. This approach reveals strengths and limitations that may not be apparent if relying solely on accuracy, especially for unbalanced datasets, which are common in biomedical data. In this study, we evaluated these multiple performance metrics to assess each classifier's performance on imbalanced, undersampled, and oversampled datasets. We aimed to identify the technique that most effectively mitigated class imbalance effects and enhanced model robustness.

Given the modest sample size and the exploratory nature of the analysis, performance comparisons across classifiers, feature sets, and balancing strategies were evaluated descriptively using cross-validated metrics to highlight general performance patterns. These observed differences provide useful preliminary indications of how different models and feature sets behave under the tested conditions.

Results

The results presented provide insight into how each feature set and sampling method influenced the performance for the T2DM versus control classification task.

Classification Performance With Individual Feature Sets Using the Full (Unbalanced) Dataset

Table 2 shows a comparison of the performance metrics for individual features using the full (unbalanced) dataset, for all classifiers (LR, SVM with linear kernel, and SVM with RBF kernel).

XSL•FO
RenderX

Table . Classification performance using unbalanced data (all samples)[a].

Feature set and classifier, mean (SD)	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[b], mean (SD)
HRV[c]						
LR[d]	0.487 (0.152)	0.553 (0.077)	0.750 (0.354)	0.100 (0.224)	0.610 (0.192)	0.525 (0.130)
SVM[e] linear	0.587 (0.084)	0.620 (0.073)	0.900 (0.224)	0.100 (0.224)	0.718 (0.098)	0.428 (0.208)
SVM RBF[f]	0.587 (0.084)	0.620 (0.073)	0.900 (0.224)	0.100 (0.224)	0.718 (0.098)	0.617 (0.311)
FRF[g]						
LR	0.613 (0.168)	0.627 (0.128)	0.800 (0.298)	0.233 (0.325)	0.692 (0.201)	0.606 (0.268)
SVM linear	0.620 (0.073)	0.620 (0.073)	0.933 (0.149)	0.067 (0.149)	0.744 (0.099)	0.533 (0.326)
SVM RBF	0.620 (0.073)	0.620 (0.073)	0.933 (0.149)	0.067 (0.149)	0.744 (0.099)	0.464 (0.137)
IR[h]						
LR	0.660 (0.106)	0.700 (0.183)	0.933 (0.149)	0.200 (0.447)	0.773 (0.060)	0.372 (0.205)
SVM linear	0.620 (0.073)	0.633 (0.075)	0.933 (0.149)	0.100 (0.224)	0.747 (0.073)	0.575 (0.205)
SVM RBF	0.620 (0.073)	0.620 (0.073)	1.000 (0.000)	0.000 (0.000)	0.763 (0.058)	0.700 (0.126)

[a]Performance metrics using individual features (HRV, FRF, or IR) for each classification model (LR, SVM with linear kernel, and SVM with RBF kernel) using the full, unbalanced dataset.

[b]AUC-ROC: area under the receiver operating characteristic curve.

[c]HRV: heart rate variability.

[d]LR: logistic regression.

[e]SVM: support vector machine.

[f]RBF: radial basis function.

[g]FRF: frequency response function.

[h]IR: impulse response.

Upon analyzing the classification performance metrics obtained for the individual feature sets—HRV, FRF, and IR metrics—we observed notable differences. Using the unbalanced dataset, the IR feature set yielded the highest metric values among the 3 domains, with accuracy (0.660, SD 0.106; LR), precision (0.700, SD 0.183; LR), recall (0.933, SD 0.149; LR and SVM linear), and F_1-score (0.773, SD 0.060; LR). FRF showed intermediate performance, while HRV consistently exhibited the lowest values across classifiers.

A comparative analysis of model performance indicated that the SVM linear and SVM RBF models performed comparably to the LR model across most metrics. For the IR feature set, the SVM RBF model achieved the highest recall (1.000, SD 0.000), while both LR and SVM linear classifiers had a recall of 0.933 (SD 0.149). The LR classifier exhibited the highest precision (0.700, SD 0.183) and F_1-score (0.773, SD 0.060).

Specificity was low across all feature sets in the unbalanced dataset, in which approximately two-thirds of the subjects are T2DM, hindering accurate classification of negative samples. This underscores the importance of using appropriate strategies to address data imbalance for improved model performance.

Impact of NearMiss-1 Undersampling on Individual Feature Sets

Tables 3 and 4 show a comparison of performance metrics for individual features using NM undersampling and SMOTE oversampling, respectively. NM improved performance across all feature sets. The IR feature often demonstrated comparatively stronger performance (eg, LR: accuracy 0.770, SD 0.179; precision 0.783, SD 0.217; recall 0.900, SD 0.224; specificity 0.633, SD 0.415; and F_1-score 0.798, SD 0.140), exceeding both HRV and FRF features in several metrics. While FRF features performed well, particularly in recall (1.000, SD 0.00) for both SVM linear and SVM RBF models, the IR features provided more balanced performance across all metrics. The HRV feature set consistently exhibited the lowest performance metrics in this scenario.

XSL•FO
RenderX

Table . Comparison of performance metrics across feature sets and classifiers using NearMiss-1 (NM) undersampling (individual features)[a].

Feature set and classifier	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[b], mean (SD)
HRV[c]						
LR[d]	0.640 (0.272)	0.567 (0.365)	0.733 (0.435)	0.567 (0.253)	0.783 (0.158)	0.767 (0.224)
SVM[e] linear	0.730 (0.192)	0.683 (0.207)	0.900 (0.224)	0.567 (0.253)	0.765 (0.190)	0.767 (0.224)
SVM RBF[f]	0.730 (0.192)	0.683 (0.207)	0.900 (0.224)	0.567 (0.253)	0.765 (0.190)	0.767 (0.224)
FRF[g]						
LR	0.560 (0.251)	0.580 (0.239)	0.767 (0.325)	0.400 (0.418)	0.628 (0.230)	0.533 (0.298)
SVM linear	0.640 (0.251)	0.630 (0.244)	1.000 (0.000)	0.300 (0.447)	0.752 (0.173)	0.667 (0.204)
SVM RBF	0.640 (0.251)	0.630 (0.244)	1.000 (0.000)	0.300 (0.447)	0.752 (0.173)	0.500 (0.373)
IR[h]						
LR	0.770 (0.179)	0.783 (0.217)	0.900 (0.224)	0.633 (0.415)	0.798 (0.140)	0.700 (0.447)
SVM linear	0.600 (0.235)	0.600 (0.235)	1.000 (0.000)	0.200 (0.447)	0.731 (0.163)	0.767 (0.253)
SVM RBF	0.540 (0.185)	0.525 (0.145)	0.700 (0.447)	0.367 (0.415)	0.654 (0.178)	0.733 (0.308)

[a]Performance metrics of balanced datasets using individual features (HRV, FRF, or IR), by applying NM undersampling.

[b]AUC-ROC: area under the receiver operating characteristic curve.

[c]HRV: heart rate variability.

[d]LR: logistic regression.

[e]SVM: support vector machine.

[f]RBF: radial basis function.

[g]FRF: frequency response function.

[h]IR: impulse response.

XSL•FO
RenderX

Table . Comparison performance metrics across feature sets and classifiers using Synthetic Minority Oversampling Technique (SMOTE)–balanced data (individual feature sets)[a].

Feature set and classifier	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[b], mean (SD)
HRV[c]						
LR[d]	0.550 (0.167)	0.619 (0.230)	0.617 (0.361)	0.500 (0.373)	0.547 (0.171)	0.654 (0.169)
SVM[e] linear	0.607 (0.164)	0.586 (0.114)	0.833 (0.236)	0.383 (0.274)	0.673 (0.130)	0.717 (0.162)
SVM RBF[f]	0.604 (0.192)	0.686 (0.288)	0.683 (0.207)	0.567 (0.435)	0.639 (0.136)	0.683 (0.190)
FRF[g]						
LR	0.446 (0.215)	0.450 (0.132)	0.633 (0.280)	0.250 (0.250)	0.523 (0.187)	0.408 (0.264)
SVM linear	0.446 (0.074)	0.380 (0.217)	0.633 (0.375)	0.250 (0.306)	0.592 (0.069)	0.300 (0.326)
SVM RBF	0.450 (0.133)	0.394 (0.229)	0.683 (0.410)	0.200 (0.209)	0.623 (0.097)	0.233 (0.320)
IR[h]						
LR	0.529 (0.156)	0.587 (0.250)	0.600 (0.379)	0.517 (0.291)	0.520 (0.190)	0.629 (0.193)
SVM linear	0.611 (0.062)	0.569 (0.041)	0.950 (0.112)	0.267 (0.181)	0.708 (0.038)	0.771 (0.062)
SVM RBF	0.532 (0.088)	0.434 (0.247)	0.617 (0.439)	0.417 (0.328)	0.615 (0.146)	0.692 (0.216)

[a]Performance metrics of balanced datasets using individual features (HRV, FRF, or IR), by applying SMOTE oversampling.

[b]AUC-ROC: area under the receiver operating characteristic curve.

[c]HRV: heart rate variability.

[d]LR: logistic regression.

[e]SVM: support vector machine.

[f]RBF: radial basis function.

[g]FRF: frequency response function.

[h]IR: impulse response.

Notably, applying NM undersampling improved specificity across all classification models (eg, LR: increased from 0.100, SD 0.224 to 0.567, SD 0.253), enhancing its reliability in classifying negative samples (ie, control subjects) while reducing false positives.

The SVM linear and SVM RBF models tended to show higher accuracy and recall than the LR classifier for the HRV and FRF feature sets, although their F_1-scores were not consistently higher. For the IR feature set, the LR classifier generally produced metrics that were equal to or slightly higher than those of the SVM models. The 2 SVM classifiers exhibited very similar performance metrics across each individual feature set, with the RBF kernel showing slight advantages in certain cases. Overall, the IR feature set paired with the LR model showed some of the comparatively stronger results on the balanced dataset.

Classification Performance Using SMOTE Oversampling on Individual Feature Sets

SMOTE oversampling enhanced performance across all feature sets compared to the unbalanced dataset. Within this setting, IR features frequently showed comparatively strong performance across classifiers, with the LR model showing the highest precision (0.587, SD 0.250) and specificity (0.517, SD 0.291) among all classifiers. The SVM linear classifier achieved the highest accuracy (0.611, SD 0.062), recall (0.950, SD 0.112), F_1-score (0.708, SD 0.038), and AUC-ROC (0.771, SD 0.062), while the highest specificity was obtained from the LR model (0.517, SD 0.291).

The HRV features produced the next highest metric values, with the SVM RBF model showing the highest precision (0.686, SD 0.288) and specificity (0.567, SD 0.435). Both SVM models showed comparable F_1-score (0.673, SD 0.130 for SVM linear) and AUC-ROC (0.717, SD 0.162 for SVM linear). While the LR model benefited from SMOTE balancing compared to the original dataset, it showed slightly lower overall metrics relative to the SVM models.

The FRF feature set, however, showed lower overall performance across all classifiers compared to HRV and IR, indicating limited discriminatory power, particularly in specificity and precision. This suggests that the FRF indices may lack comprehensive information required for effective classification across all classifiers when using the SMOTE-oversampled dataset.

In terms of model comparisons, the SVM linear classifier generally showed comparatively strong performance across feature sets, particularly when paired with IR metrics. The SVM RBF model demonstrated high recall but tended to have lower precision and F_1-scores than SVM linear. Although SMOTE improved LR performance, it lagged behind SVM models,

XSL•FO
RenderX

except when using the IR feature set, where it showed competitive results.

Overall, SMOTE balancing improved specificity and overall classification reliability over the unbalanced dataset. The combination of IR features with the SVM linear model produced some of the strongest performance patterns observed in this analysis, underscoring how feature-classifier interactions can influence discrimination. These patterns offer hypothesis-generating observations that warrant evaluation in larger datasets.

Comparative Effectiveness of Nearmiss-1 and SMOTE for Classifier Performance on Individual Feature Sets

Comparing the performance metrics of the NM and SMOTE data balancing approaches, both techniques clearly enhanced classification metrics over the unbalanced dataset but had different strengths. NM undersampling often led to higher specificity across classifiers, indicating better identification of negative samples. This was particularly notable with the IR feature set, where the LR (NM) model achieved higher recall (0.900, SD 0.224) and a superior F_1-score (0.798, SD 0.140) compared to LR (SMOTE), which showed lower recall (0.600, SD 0.379) and F_1-score (0.520, SD 0.19).

The recall and F_1-scores of the SVM linear model using balanced data from either NM or SMOTE strategies were mostly similar, suggesting comparable performance of the two balancing methods for this classifier. While SVM RBF (NM) showed slightly better recall and F_1-score than SVM RBF (SMOTE), the differences were marginal due to the high variability in the results.

In general, SMOTE effectively enhanced sensitivity, well-suited for identifying T2DM cases, though its effectiveness varied by classifier and feature set, with inconsistent AUC-ROC gains.

NM, on the other hand, was more effective for improving specificity and F_1-scores (eg, LR and SVM linear with IR), particularly advantageous for non-T2DM classification. These results underscore the importance of selecting the appropriate data balancing approach based on specific classification goals and the clinical implications of false positives versus false negatives.

Combined Feature Sets Analysis

We also evaluated whether combining HRV, FRF, and IR features would enhance classification performance. Table 5 shows a comparison of the performance metrics for the combined HRV+FRF and HRV+IR feature sets, along with the metrics of the individual IR feature set, using the full (unbalanced) dataset, for all classifiers (LR, SVM linear, and SVM RBF).

For the full, unbalanced dataset, combinations (HRV+FRF and HRV+IR) did not provide a significant advantage over using individual FRF or IR feature sets. The IR feature set alone generally showed the strongest performance patterns across models, with comparatively higher accuracy, precision, recall, and F_1-score relative to the combined feature sets in this setting.

Tables 6 and 7 show a comparison of performance metrics for the combined HRV+FRF and HRV+IR feature sets, along with those for the individual IR feature set, using NM undersampling and SMOTE oversampling, respectively.

XSL•FO
RenderX

Table . Comparison of performance metrics for combined feature sets (HRV[a]+FRF[b], HRV+IR[c]) and individual lR feature set using unbalanced data[d].

Feature set and classifier	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[e], mean (SD)
HRV+FRF						
LR[f]	0.553 (0.141)	0.520 (0.292)	0.700 (0.447)	0.233 (0.325)	0.730 (0.109)	0.469 (0.291)
SVM[g] linear	0.620 (0.182)	0.613 (0.173)	0.867 (0.298)	0.167 (0.236)	0.714 (0.219)	0.422 (0.248)
SVM RBF[h]	0.547 (0.166)	0.567 (0.149)	0.800 (0.298)	0.067 (0.149)	0.661 (0.208)	0.656 (0.374)
HRV+IR						
LR	0.553 (0.141)	0.600 (0.091)	0.833 (0.236)	0.100 (0.224)	0.687 (0.124)	0.467 (0.302)
SVM linear	0.587 (0.084)	0.620 (0.073)	0.883 (0.162)	0.100 (0.224)	0.720 (0.073)	0.492 (0.298)
SVM RBF	0.547 (0.117)	0.587 (0.084)	0.883 (0.162)	0.000 (0.000)	0.701 (0.098)	0.597 (0.284)
IR						
LR	0.660 (0.106)	0.700 (0.183)	0.933 (0.149)	0.200 (0.447)	0.773 (0.060)	0.372 (0.205)
SVM linear	0.620 (0.073)	0.633 (0.075)	0.933 (0.149)	0.100 (0.224)	0.747 (0.073)	0.575 (0.205)
SVM RBF	0.620 (0.073)	0.620 (0.073)	1.000 (0.000)	0.000 (0.000)	0.763 (0.058)	0.700 (0.126)

[a]HRV: heart rate variability.

[b]FRF: frequency response function.

[c]IR: impulse response.

[d]Performance metrics using combined feature sets (HRV+FRF and HRV+IR), compared to the individual IR feature set, for each classification model (LR, SVM linear, and SVM RBF) using the full, unbalanced dataset.

[e]AUC-ROC: area under the receiver operating characteristic curve.

[f]LR: logistic regression.

[g]SVM: support vector machine.

[h]RBF: radial basis function.

XSL•FO
RenderX

Table . Comparison of performance metrics for combined feature sets (HRV[a]+FRF[b], HRV+IR[c]) and individual IR set, using NearMiss-1 (NM) balanced data[d].

Feature set and classifier	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[e], mean (SD)
HRV+FRF						
LR[f]	0.680 (0.280)	0.667 (0.236)	0.833 (0.236)	0.533 (0.361)	0.733 (0.221)	0.667 (0.295)
SVM[g] linear	0.730 (0.076)	0.667 (0.000)	0.933 (0.149)	0.533 (0.075)	0.773 (0.060)	0.767 (0.253)
SVM RBF[h]	0.830 (0.172)	0.800 (0.183)	0.933 (0.149)	0.733 (0.253)	0.853 (0.145)	0.800 (0.274)
HRV+IR						
LR	0.680 (0.344)	0.767 (0.325)	0.667 (0.312)	0.700 (0.447)	0.693 (0.300)	0.783 (0.298)
SVM linear	0.730 (0.192)	0.767 (0.224)	0.833 (0.236)	0.633 (0.415)	0.760 (0.146)	0.800 (0.274)
SVM RBF	0.680 (0.125)	0.700 (0.183)	0.833 (0.236)	0.533 (0.361)	0.720 (0.073)	0.667 (0.312)
IR						
LR	0.770 (0.179)	0.783 (0.217)	0.900 (0.224)	0.633 (0.415)	0.798 (0.140)	0.700 (0.447)
SVM linear	0.600 (0.235)	0.600 (0.235)	1.000 (0.000)	0.200 (0.447)	0.731 (0.163)	0.767 (0.253)
SVM RBF	0.540 (0.185)	0.525 (0.145)	0.700 (0.447)	0.367 (0.415)	0.654 (0.178)	0.733 (0.308)

[a]HRV: heart rate variability.

[b]FRF: frequency response function.

[c]IR: impulse response.

[d]Performance metrics using combined feature sets (HRV+FRF and HRV+IR), compared to the individual IR feature set, for each classification model (LR, SVM linear, and SVM RBF) using NM undersampling.

[e]AUC-ROC: area under the receiver operating characteristic curve.

[f]LR: logistic regression.

[g]SVM: support vector machine.

[h]RBF: radial basis function.

XSL•FO
RenderX

Table . Comparison of performance metrics for combined feature sets (HRV[a]+FRF[b], HRV+IR[c]) and individual IR set, using Synthetic Minority Oversampling Technique (SMOTE)–balanced data[d].

Feature set and classifier	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	Specificity, mean (SD)	F_1-score, mean (SD)	AUC-ROC[e], mean (SD)
HRV+FRF						
LR[f]	0.421 (0.183)	0.422 (0.137)	0.550 (0.411)	0.350 (0.253)	0.444 (0.220)	0.446 (0.156)
SVM[g] linear	0.586 (0.078)	0.553 (0.077)	0.900 (0.137)	0.283 (0.046)	0.680 (0.073)	0.567 (0.231)
SVM RBF[h]	0.586 (0.192)	0.583 (0.373)	0.483 (0.291)	0.683 (0.207)	0.646 (0.105)	0.600 (0.279)
HRV+IR						
LR	0.671 (0.139)	0.650 (0.137)	0.733 (0.181)	0.617 (0.112)	0.686 (0.146)	0.688 (0.201)
SVM linear	0.589 (0.200)	0.581 (0.180)	0.900 (0.137)	0.317 (0.335)	0.690 (0.135)	0.729 (0.116)
SVM RBF	0.700 (0.128)	0.783 (0.217)	0.683 (0.207)	0.750 (0.306)	0.691 (0.097)	0.742 (0.139)
IR						
LR	0.529 (0.156)	0.587 (0.250)	0.600 (0.379)	0.517 (0.291)	0.520 (0.190)	0.629 (0.193)
SVM linear	0.611 (0.062)	0.569 (0.041)	0.950 (0.112)	0.267 (0.181)	0.708 (0.038)	0.771 (0.062)
SVM RBF	0.532 (0.088)	0.434 (0.247)	0.617 (0.439)	0.417 (0.328)	0.615 (0.146)	0.692 (0.216)

[a]HRV: heart rate variability.

[b]FRF: frequency response function.

[c]IR: impulse response.

[d]Performance metrics using combined feature sets (HRV+FRF and HRV+IR), compared to the individual IR feature set, for each classification model (LR, SVM linear, and SVM RBF) using SMOTE oversampling.

[e]AUC-ROC: area under the receiver operating characteristic curve.

[f]LR: logistic regression.

[g]SVM: support vector machine.

[h]RBF: radial basis function.

In NM balanced datasets, the combined HRV+FRF feature set yielded higher values than the individual IR feature set in accuracy (0.830, SD 0.172 vs 0.770, SD 0.179), precision (0.800, SD 0.183 vs 0.783, SD 0.217), and F_1-score (0.853, SD 0.145 vs 0.798, SD 0.140) for the SVM RBF classifier. HRV+FRF also produced higher metric values than HRV alone across all metrics for SVM RBF (accuracy: 0.830, SD 0.172 vs 0.730, SD 0.192; precision: 0.800, SD 0.183 vs 0.683, SD 0.207; and F_1-score: 0.853, SD 0.145 vs 0.765, SD 0.190). Among classifiers, SVM RBF exhibited some of the strongest observed performance with HRV+FRF, while SVM linear showed comparable patterns with both HRV+FRF and HRV+IR.

For SMOTE-processed data, the HRV+IR combined feature set yielded higher values than the individual IR feature set in most metrics, particularly in accuracy (SVM RBF: 0.700, SD 0.128 vs 0.532, SD 0.088), precision (0.783, SD 0.217 vs 0.434, SD 0.247), and AUC-ROC (0.742, SD 0.139 vs 0.692, SD 0.216). However, IR alone retained slightly better recall (0.950, SD 0.112 vs 0.900, SD 0.137) and F_1-score (0.708, SD 0.038 vs 0.690, SD 0.135) with the SVM linear classifier. Within this SMOTE-balanced setting, the SVM RBF model showed some of the strongest performance patterns when paired with HRV+IR, while SVM linear also performed well, maintaining a high recall (0.900, SD 0.137) and a comparable F_1-score (0.690, SD 0.135), indicating effective classification sensitivity. The LR model produced moderate metrics with HRV+IR (accuracy: 0.671, SD 0.139; precision: 0.650, SD 0.137; AUC-ROC: 0.688, SD 0.201), highlighting the use of this combined feature set even for simpler models.

In summary, the combined feature sets improved overall classification performance in the SMOTE-balanced dataset, and HRV+IR generally produced comparatively strong results across models. Among the classifiers evaluated, SVM RBF tended to show some of the higher metric values when paired with this feature set.

Summary of Performance Findings

Across classifiers and balancing conditions, the IR feature set often showed comparatively higher performance as an individual feature set, particularly in terms of recall and F_1-score with SVM linear. Combining IR with HRV offered performance improvements in the SMOTE-balanced scenario and more modest gains with NM. In general, combined feature sets tended to show enhanced performance relative to the unbalanced datasets, particularly when SMOTE was used.

SMOTE tended to increase recall and overall sensitivity, especially for SVM linear, whereas NM produced higher specificity and F_1-scores. These patterns suggest that the choice of sampling technique should be guided by the classification goals and the clinical importance of false positives versus false negatives.

XSL•FO
RenderX

No feature set consistently outperformed others across all metrics or balancing strategies. For instance, the combined HRV+FRF feature set performed well with NM, while HRV+IR showed comparatively better performance under SMOTE. The IR feature set remained competitive as a standalone option, showing similar sensitivity-based performance to HRV+FRF. Under NM, HRV+FRF achieved slightly higher accuracy, specificity, F_1-score, and AUC-ROC than IR alone, while under SMOTE, HRV+IR showed improvements in accuracy (0.168), precision (0.349), specificity (0.333), and AUC-ROC (0.050) relative to IR. IR alone retained slightly higher recall (0.050) and F_1-score (0.018) in that setting, without added feature complexity.

Although combined feature sets offered incremental benefits in several cases, selecting among them should weigh these gains against increased model complexity and limited sample size. For applications where accuracy and precision are emphasized, the HRV+IR feature set under SMOTE may warrant further investigation. For settings prioritizing simplicity or sensitivity to potential T2DM cases, the IR feature set under NM may remain a practical alternative. Overall, these findings represent hypothesis-generating patterns that may guide future analysis in larger and more diverse cohorts.

Discussion

Overview of Findings

This study investigated the effectiveness of various feature sets, classification models, and data balancing techniques for distinguishing individuals with and without T2DM. The findings highlight the strengths and limitations of different approaches while examining the discriminative value of HRV, FRF, and IR metrics. To the best of our knowledge, this is the first study to evaluate these complementary domains of autonomic and cardiorespiratory regulation within a ML framework in the context of T2DM. By systematically assessing individual and combined feature sets, this study provides exploratory insight into physiologically grounded patterns that may differentiate cardiorespiratory regulation between individuals with and without T2DM, supporting future investigations in larger and independently validated cohorts.

Principal Results

Classification Performance Across Feature Sets

Across individual feature sets, IR function metrics tended to show comparatively higher predictive performance for distinguishing individuals with and without T2DM. This pattern may reflect IR metric's ability to capture causal, time-domain characteristics of RCC, which could be sensitive to subtle regulatory differences associated with diabetes. Prior work in related fields, such as obstructive sleep apnea [21], has demonstrated the value of causal analyses of cardiorespiratory interactions, and the present findings extend that insight in the context of diabetes. To the authors' knowledge, this is the first study to evaluate IR metrics in this setting, providing preliminary evidence that motivates further validation in larger and clinically characterized cohorts.

Similarly, Marmarelis et al [39] used IR estimation methods to model the causal, directional influences of arterial blood pressure and CO_2 fluctuations (inputs) on cerebral blood velocity (CBV, output). In particular, from the estimated IRs, they derived principal dynamic modes—a data-based modeling technique that decomposes these responses into key dynamic components—identifying significant reductions in principal dynamic modes gain (indicating weakened regulatory responses) for both arterial blood pressure-to-CBV and CO_2-to-CBV pathways in patients with T2DM compared to controls. This approach enabled the creation of a composite diagnostic index with an AUC of 0.78 for differentiating T2DM from controls, underscoring the value of integrating IR-based directional modeling and noncausal measures in detecting subtle physiological impairments.

In our study, when feature sets were combined (HRV+FRF or HRV+IR), we observed improved or comparable performance compared to IR alone. These patterns likely reflect the complementary physiological information captured by the different domains: HRV summarizes the overall frequency-domain structure of cardiac variability, FRF metrics describe frequency-specific transfer properties of respiratory–cardiac interactions, and IR metrics quantify causal dynamic responsiveness to respiratory inputs.

Prior work also supports the integrative use of multivariate autonomic and cardiorespiratory descriptors. Emerging evidence suggests that metrics targeting specific physiological pathways can reveal regulatory differences not captured by global HRV indices. For example, reductions in respiratory–cardiac interactions have been suggested as early indicators of impaired autonomic regulation in type 2 diabetes [40]. Similarly, baroreflex measures derived from causal, model-based approaches have outperformed traditional spontaneous indices in predicting clinical outcomes and identifying autonomic impairment in patient cohorts [41]. Together, these findings highlight the value of model-based approaches for characterizing pathway-specific physiological regulation.

Taken together, these findings suggest that HRV, FRF, and IR metrics probe different aspects of cardiorespiratory autonomic regulation and may offer complementary perspectives when assessed individually and in combination. Nonetheless, given the exploratory nature of the present analysis and the modest sample size, improvements with combined feature sets may also partly reflect increased feature dimensionality rather than purely additive physiological contributions.

It is important to emphasize that the classifiers in this study distinguish diabetes status, not clinically diagnosed autonomic dysfunction. Therefore, the observed differences likely reflect diabetes-related physiological alterations that may involve autonomic components, but our models cannot be interpreted as detecting or predicting autonomic impairment at the individual level.

Influence of Balancing Techniques on Performance

Balancing techniques significantly influenced classification performance. NM consistently improved specificity and F_1-scores, refining class distinction by retaining T2DM (the

XSL•FO
RenderX

majority class) instances closest to the opposite class. This is particularly important for imbalanced datasets where negative class identification is often challenging.

In contrast, SMOTE enhanced sensitivity (recall) for some classifiers (especially SVM linear), by generating synthetic samples for the minority class (controls), though specificity gains were less consistent. SMOTE's synthetic samples are not tailored to emphasize the decision boundary and may lead to overlap between classes and, consequently, a reduced precision in distinguishing control cases.

Similarly, in a study to assess the efficacy of different ML models and balancing techniques for diabetes diagnosis using an imbalanced multiclass dataset (with class 0: nondiabetic, class 1: prediabetic as minorities, and class 2: diabetic as the majority) [42], the authors found that the overall recall (macro-averaged across classes) for SVM (linear) and SVM (RBF) improved with SMOTE oversampling compared to NM undersampling, though per-class results were mixed (eg, lower class 1 recall for SVM RBF with SMOTE, despite gains in classes 0 and 2). In contrast, when LR was used as the classifier, recall was substantially higher using NM balancing compared to SMOTE, with NM outperforming SMOTE across all classes—substantial gains in class 1 (prediabetic, the minority class, +0.40) and smaller improvements in class 0 (+0.14) and class 2 (diabetic, the majority class, +0.05)—suggesting NM is more effective at balancing the dataset for this classifier, particularly for the minority class. Specificity was not evaluated in this study.

These conceptual differences highlight the importance of selecting a balancing technique aligned with the predictive goals. The choice depends on prioritizing specificity (NM) or sensitivity (SMOTE), guided by the clinical implications of false positives versus false negatives. In clinical screening contexts, where the aim is often to identify individuals at higher metabolic or cardiovascular risk, minimizing false negatives (maximizing recall/sensitivity) is typically more important, as missed cases may delay further evaluation and preventive care. Because SMOTE tended to improve recall across feature sets and classifiers, it may be better aligned with population-level screening or risk-stratification workflows where sensitivity is prioritized.

Classifier-Specific Observations

While more complex models like SVM RBF occasionally achieved higher performance, linear SVM and LR models offered comparable results, especially when paired with IR features. This suggests that even simpler, more interpretable models can perform competitively when the feature set is physiologically meaningful—an important consideration for clinical adoption.

Implications for Understanding T2DM Classification

The findings of this study suggest that dynamic, causal features, particularly those derived from IR metrics, may capture physiologically meaningful differences in cardiorespiratory regulation between individuals with and without T2DM. IR measures quantify the responsiveness of RRI to respiratory perturbations, which may reflect aspects of autonomic adaptability not fully represented by traditional HRV and FRF metrics. In this context, reductions in IR measures may indicate changes in RCC that warrant further investigation in larger cohorts.

Across classifiers and balancing strategies, the generally strong performance of the IR feature set indicates that modeling causal, time-domain dynamics can provide useful discriminatory information when exploring how T2DM relates to autonomic and cardiorespiratory regulation. The integration of static (HRV and FRF) and dynamic (IR) metrics offers a preliminary multivariate perspective on physiological regulation in T2DM, supporting the hypothesis that complementary domains may capture different aspects of diabetes-related physiological differences.

Because this analysis is exploratory and based on a modest sample size, these interpretations should be viewed as hypothesis-generating. Future studies with well-defined autonomic phenotyping will be necessary to clarify the extent to which these physiological patterns reflect autonomic regulation, microvascular alterations, metabolic factors, or other T2DM-related mechanisms.

Limitations

While this study provides valuable insights into the classification of T2DM using various feature sets, classifiers, and data balancing techniques, several limitations should be noted.

A primary limitation is the relatively modest sample size of the PhysioNet datasets used, particularly for ML applications in which multiple feature sets and classifiers are evaluated. This limited sample size reduces statistical power, increases fold-to-fold variability in cross-validation, and heightens the risk of overfitting—despite our use of feature standardization, correlation-based feature reduction, and 5-fold cross-validation to mitigate these issues. Consequently, the generalizability of the findings to broader and more heterogeneous T2DM populations is uncertain. Larger, more diverse datasets will be essential to validate the models, confirm the stability of the feature sets, and establish their applicability across different demographic and clinical subgroups. For these reasons, the conclusions of the present study should be interpreted as preliminary and exploratory.

Protocol-related differences across datasets represent an additional limitation. While the protocol for the Cerebral Perfusion and Cognitive Decline in Type 2 Diabetes dataset [17] included paced breathing, the Cerebral Vasoregulation in Diabetes dataset [16] did not involve controlled respiratory conditions. These differences may influence respiratory patterns, autonomic engagement, and cardiorespiratory coupling dynamics, potentially affecting the FRF and IR estimates used in this study. As such, part of the observed variability may reflect protocol-specific physiological factors rather than group differences alone. Future studies using harmonized experimental designs will be important to isolate and interpret these effects.

A further methodological limitation relates to the implementation of the class-balancing procedures. In this exploratory analysis, NM undersampling and SMOTE oversampling were applied to the full usable dataset prior to

XSL•FO
RenderX

generating the 5-fold cross-validation partitions, rather than separately within each training fold. This approach enabled direct comparison of balancing strategies under fixed-class distributions, but it also means that, particularly for SMOTE, synthetic samples were generated using neighborhood information from the entire dataset and could subsequently appear in both training and test folds. This introduces a degree of information leakage and may partially contribute to the variability observed in some performance metrics. Such variability is also expected given the modest sample size, where each fold contains relatively few test samples, making sensitivity, specificity, and F_1-score more sensitive to fold composition. For these reasons, the performance estimates should be viewed as preliminary. Future studies with larger datasets will be able to implement fold-wise balancing and preprocessing to avoid this issue and obtain more stable and generalizable results.

A related limitation concerns the correlation-based feature filtering step. To maintain consistent feature definitions across all model configurations, correlation filtering (threshold =0.8) was applied once to the full usable dataset rather than separately within each cross-validation fold. This choice avoided the instability and inconsistency that fold-wise feature selection would likely introduce in a small dataset, but it also means that correlation structure from the entire dataset—including samples later assigned to the test folds—contributed to the filtering process. As a result, this preprocessing decision introduces a potential source of information leakage and could lead to mildly optimistic performance estimates. Future work with larger cohorts will enable fold-wise filtering and more sophisticated assessments of feature redundancy (eg, variance inflation factor, mutual information, principal component analysis–based methods) while preserving model stability.

A key challenge of incorporating IR metrics is the need for respiration measurements, which introduces an extra layer of complexity to the practical implementation of the clinical testing protocol. This additional signal channel can complicate data acquisition and processing, potentially limiting the feasibility of the approach in resource-constrained settings. However, recent technological advancements have mitigated some of these challenges by introducing standalone instrumentation systems that are both portable and low-cost, capable of simultaneously measuring ECG and respiration signals. Examples of such systems include the Protocentral tinyECG module (Protocentral), which uses the MAX30001 chip (Maxim Integrated/Analog Devices) to integrate biopotential and bioimpedance channels for ECG and respiration measurements, and the Equivital LifeMonitor (Equivital), which provides clinical-grade ECG and breathing rate measurements via impedance. These systems offer user-friendly interfaces and affordability, potentially reducing barriers to adoption in clinical practice.

It should be acknowledged that BMI is a known modulator of autonomic regulation. Because the usable dataset was small and BMI was strongly collinear with T2DM status, adjusting for BMI (eg, via analysis of covariance or covariate-adjusted modeling) would have further reduced statistical power and produced potentially unstable estimates. The goal of this analysis was to examine the mechanistic and discriminative value of the physiological features (HRV, FRF, and IR) rather than to isolate covariate-adjusted effects. Therefore, BMI was not included as a covariate. We acknowledge that part of the observed group differences may reflect the physiological influence of adiposity rather than diabetes-specific autonomic dysfunction alone. Future studies with larger and more diverse cohorts will be needed to disentangle the independent contributions of adiposity and diabetes.

Moreover, because HRV, FRF, and IR feature domains are all derived from cardiac timing data—and FRF and IR additionally incorporate respiratory inputs—some degree of redundancy among features is expected. To reduce multicollinearity, we applied correlation-based filtering (threshold =0.8) across all feature configurations, including the standalone and combined feature sets. Although HRV, FRF, and IR capture different but related aspects of autonomic and cardiorespiratory regulation, residual redundancy may remain, and some performance gains from combined feature sets may partly reflect increased dimensionality rather than strictly complementary physiological information. Larger datasets will be needed to more clearly differentiate the unique versus overlapping contributions of these domains.

Additionally, the study's reliance on specific classifiers (LR, SVM linear, and SVM RBF) and balancing techniques (NM and SMOTE) may restrict generalizability to other ML frameworks or preprocessing pipelines. Furthermore, while combined feature sets occasionally improved performance, these gains were modest relative to the added model complexity. Future studies should evaluate the trade-offs between feature aggregation, interpretability, and computational efficiency.

Comparison With Prior Work

As stated by the American Diabetes Association [43], CAN is asymptomatic in its early stages and detected only by HRV calculated from recording an ECG either during a shift from a seated to a standing posture or during a 1 2 minute deep breathing test in the doctor's office, both of which require patient cooperation. Using time, frequency, and nonlinear HRV indices from both resting and orthostatic challenge data, Rathod et al [44] showed that a classification and regression tree model showed an accuracy of 0.840, sensitivity of 0.895, a specificity of 0.667, and an AUC of 0.78 compared to resting HRV alone with 0.751 accuracy, 0.864 sensitivity, 0.392 specificity, with an AUC of 0.63 for differentiating autonomic dysfunction in nondiabetic control and T2DM.

In our study, the HRV feature set obtained from sitting data showed comparatively lower performance for distinguishing individuals with and without T2DM, also suggesting that sitting HRV alone may lack the granularity needed to capture physiological differences associated with diabetes. IR metrics, which capture the dynamic influence of respiration on RRI, frequently demonstrated stronger discriminative performance than HRV and FRF features within our dataset. For example, the IR feature set obtained from sitting data showed an accuracy of 0.770, sensitivity of 0.900, specificity of 0.633, and an AUC of 0.700 (LR using NM balancing) for differentiating T2DM from controls, values comparable to those reported by Rathod

XSL•FO
RenderX

et al [44] despite using resting data alone. These findings suggest that causal, time-domain representations of respiratory-cardiac interactions may capture physiological distinctions between T2DM and control groups that are not fully reflected in resting HRV or noncausal FRF metrics.

Our modeling framework also differs from Rathod et al [44] in its use of LR and SVM classifiers rather than a CART model. CART provides explicit, rule-based decision pathways that are easily interpretable in clinical settings. LR offers coefficient-based interpretability, whereas SVMs rely on margin-based discrimination that emphasizes classification boundaries rather than direct feature-level explanations. Therefore, these models represent alternative analytical strategies, each with distinct strengths in terms of transparency and decision structure.

Finally, although IR- and FRF-based metrics may provide physiologically motivated insights into respiratory–cardiac regulation, further work incorporating validated autonomic outcomes will be required to determine their relevance in assessing autonomic impairment in diabetes.

While our approach leverages physiologically interpretable IR metrics for passive monitoring, other studies have explored automated diabetes detection using deep learning models applied to ECG-derived signals, though often at the cost of interpretability. Swapna et al [45] used a hybrid deep neural network combining a convolutional neural network–long short-term memory (CNN–LSTM) using RRI time series (derived from ECG signals) as input, achieving 95.1% accuracy in diabetes detection. These RRIs represent raw HRV data without specific feature extraction. In a subsequent study, Swapna et al [46] integrated an SVM classifier following the CNN–LSTM architecture, improving accuracy to 95.7%.

However, the absence of additional performance metrics, such as precision, sensitivity, and specificity, obscures the models' ability to minimize false positives and false negatives. Furthermore, deep learning approaches like CNN–LSTM are noninterpretable, offering limited insight into which HRV features drive classification outcomes. In contrast, our study examined feature domains—spectral HRV, FRF, and IR metrics—chosen for their physiological grounding in autonomic and cardiorespiratory regulation. These features provide mechanistic insight by characterizing overall variability (HRV), frequency-domain transfer properties (FRF), and causal dynamic responsiveness to respiratory inputs (IR). However, while IR-derived measures such as IR magnitude, DG, and tchar are physiologically interpretable within a systems-modeling framework, they require specialized technical understanding. The value of these metrics lies in their potential to complement traditional HRV-based assessments by probing different regulatory pathways. Future work incorporating validated autonomic phenotyping will be necessary to determine whether these physiologically motivated descriptors can be translated into clinically interpretable or actionable tools.

Conclusion

This study highlights the potential value of dynamic cardiorespiratory metrics—particularly IR features—for distinguishing individuals with and without T2DM. By modeling the causal, time-domain characteristics of RCC, IR metrics frequently demonstrated comparatively strong and physiologically interpretable discriminative performance, complementing the information provided by traditional HRV and noncausal FRF measures.

Given the modest sample size and the exploratory nature of the analysis, these findings should be interpreted as preliminary. Performance estimates may be affected by dataset-specific characteristics, limited statistical power, and the risk of overfitting, and therefore may not generalize to broader populations.

Taken together, the results suggest that systems-based cardiorespiratory features—spanning variability measures, frequency-domain transfer properties, and causal dynamic responses—capture physiological differences associated with T2DM that merit further investigation. Future studies using larger and more diverse cohorts with validated autonomic phenotyping will be essential to clarify how these features relate to clinically meaningful autonomic regulation and to evaluate their broader translational relevance. Such work will help determine the extent to which these physiological domains contribute to our understanding of diabetes-related regulatory changes.

Acknowledgments

SMC would like to acknowledge the opportunity of participating as a volunteer junior researcher in the ProIC/DGP/UnB-PIBIC 2021/2022 University of Brasilia's junior research program (Programa de Iniciação Científica da Universidade de Brasilia—ProIC/DPG/UnB).

Funding

The authors would like to acknowledge the School of Technology (Faculdade de Tecnologia) of the University of Brasilia (FT/UnB) for financial support in covering the article processing fee through Edital FTD Nº 012/2025.

Data Availability

The dataset analyzed in the study is publicly available online on the PhysioNet website. The software used for the preprocessing of all datasets (Cardiorespiratory System Identification Lab [19]), as well as all univariate and multivariate analyses performed, is freely available at the corresponding author's website [47].

XSL•FO
RenderX

Authors' Contributions

Conceptualization: FMGSAO, MCKK
Formal analysis: SMC
Methodology: FMGSAO, MCKK
Resources: FMGSAO
Software: SMC
Supervision: FMGSAO, MCKK
Validation: SMC
Visualization: FMGSAO
Writing – original draft: FMGSAO, SMC
Writing – review & editing: FMGSA, SMC, MCKK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequency response function and impulse response methodologies.
[DOC File, 1012 KB - diabetes_v11i1e82084_app1.doc]

References

1. IDF Diabetes Atlas, 11th edition: International Diabetes Federation; 2025. URL: https://diabetesatlas.org/media/uploads/sites/3/2025/04/IDF_Atlas_11th_Edition_2025-1.pdf [accessed 2026-02-03]
2. Emerging Risk Factors Collaboration, Sarwar N, Gao P, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010 Jun 26;375(9733):2215-2222. [doi: 10.1016/S0140-6736(10)60484-9] [Medline: 20609967]
3. Maser RE, Mitchell BD, Vinik AI, Freeman R. The association between cardiovascular autonomic neuropathy and mortality in individuals with diabetes: a meta-analysis. Diabetes Care 2003 Jun;26(6):1895-1901. [doi: 10.2337/diacare.26.6.1895] [Medline: 12766130]
4. Eleftheriadou A, Spallone V, Tahrani AA, Alam U. Cardiovascular autonomic neuropathy in diabetes: an update with a focus on management. Diabetologia 2024 Dec;67(12):2611-2625. [doi: 10.1007/s00125-024-06242-0] [Medline: 39120767]
5. Sörensen BM, Houben A, Berendschot T, et al. Prediabetes and type 2 diabetes are associated with generalized microvascular dysfunction: the Maastricht Study. Circulation 2016 Nov 1;134(18):1339-1352. [doi: 10.1161/CIRCULATIONAHA.116.023446] [Medline: 27678264]
6. Azulay N, Olsen RB, Nielsen CS, et al. Reduced heart rate variability is related to the number of metabolic syndrome components and manifest diabetes in the sixth Tromsø study 2007-2008. Sci Rep 2022 Jul 14;12(1):11998. [doi: 10.1038/s41598-022-15824-0] [Medline: 35835836]
7. Benichou T, Pereira B, Mermillod M, et al. Heart rate variability in type 2 diabetes mellitus: a systematic review and meta-analysis. PLoS One 2018;13(4):e0195166. [doi: 10.1371/journal.pone.0195166] [Medline: 29608603]
8. Wang K, Ahmadizar F, Geurts S, et al. Heart rate variability and incident type 2 diabetes in general population. J Clin Endocrinol Metab 2023 Sep 18;108(10):2510-2516. [doi: 10.1210/clinem/dgad200] [Medline: 37022971]
9. Coopmans C, Zhou TL, Henry RMA, et al. Both prediabetes and type 2 diabetes are associated with lower heart rate variability: the Maastricht Study. Diabetes Care 2020 May;43(5):1126-1133. [doi: 10.2337/dc19-2367] [Medline: 32161051]
10. Atala YB, De Matos MR, Zantut-Wittmann DE, et al. Cardiovascular autonomic reflex tests and 7 heart rate variability indices for early diagnosis of cardiovascular autonomic neuropathy in type 2 diabetes individuals. Curr Diabetes Rev 2022;18(4):e270821195908. [doi: 10.2174/1573399817666210827130339] [Medline: 34455961]
11. Khoo MC, Belozeroff V, Berry RB, Sassoon CS. Cardiac autonomic control in obstructive sleep apnea: effects of long-term CPAP therapy. Am J Respir Crit Care Med 2001 Sep 1;164(5):807-812. [doi: 10.1164/ajrccm.164.5.2010124] [Medline: 11549537]
12. Vaschillo EG, Vaschillo B. Transfer function of the heart rate control system with respiratory input: the classical engineering approach. Presented at: Proceedings of the International Conference on Bio-inspired Systems and Signal Processing; Jan 14-17, 2009; Porto, Portugal p. 233-238. [doi: 10.5220/0001434002330238]
13. Saul JP, Berger RD, Albrecht P, Stein SP, Chen MH, Cohen RJ. Transfer function analysis of the circulation: unique insights into cardiovascular regulation. Am J Physiol 1991 Oct;261(4 Pt 2):H1231-H1245. [doi: 10.1152/ajpheart.1991.261.4.H1231] [Medline: 1928405]
14. Chon KH, Mukkamala R, Toska K, Mullen TJ, Armoundas AA, Cohen RJ. Linear and nonlinear system identification of autonomic heart-rate modulation. IEEE Eng Med Biol Mag 1997;16(5):96-105. [doi: 10.1109/51.620500] [Medline: 9313086]

XSL•FO
RenderX

15. Khoo MCK. Physiological Control Systems: Analysis, Simulation, and Estimation, 2nd edition: John Wiley & Sons; 2018. URL: https://www.wiley-vch.de/en/areas-interest/engineering/physiological-control-systems-978-1-119-05533-4 [accessed 2026-02-03]
16. Novak V, Mendez L. Cerebral vasoregulation in diabetes. PhysioNet. 2020. URL: https://physionet.org/content/cerebral-vasoreg-diabetes/1.0.0/ [accessed 2026-02-03]
17. Novak V, Quispe R, Saunders C. Cerebral perfusion and cognitive decline in type 2 diabetes. PhysioNet. 2022. URL: https://physionet.org/content/cerebral-perfusion-diabetes/1.0.1/ [accessed 2026-02-03]
18. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 1996;93(5):1043-1065. [doi: 10.1161/01.CIR.93.5.1043] [Medline: 8598068]
19. da Silva LSCB, Oliveira FMGS. CRSIDLab: a toolbox for multivariate autonomic nervous system analysis using cardiorespiratory identification. IEEE J Biomed Health Inform 2020 Mar;24(3):728-734. [doi: 10.1109/JBHI.2019.2914211] [Medline: 31056529]
20. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;32(3):230-236. [doi: 10.1109/TBME.1985.325532] [Medline: 3997178]
21. Belozeroff V, Berry RB, Khoo MCK. Model-based assessment of autonomic control in obstructive sleep apnea syndrome. Sleep 2003 Feb 1;26(1):65-73. [doi: 10.1093/sleep/26.1.65] [Medline: 12627735]
22. Saul JP. Beat-to-beat variations of heart rate reflect modulation of cardiac autonomic outflow. Physiology 1990 Feb;5(1):32-37. [doi: 10.1152/physiologyonline.1990.5.1.32]
23. Bonnemeier H. The never-ending story: unraveling the mysteries of the low-frequency component of heart rate variability. Heart Rhythm 2007 Dec;4(12):1530-1531. [doi: 10.1016/j.hrthm.2007.08.005] [Medline: 17977802]
24. Akselrod S, Gordon D, Madwed JB, Snidman NC, Shannon DC, Cohen RJ. Hemodynamic regulation: investigation by spectral analysis. Am J Physiol 1985 Oct;249(4 Pt 2):H867-H875. [doi: 10.1152/ajpheart.1985.249.4.H867] [Medline: 4051021]
25. Pagani M, Lombardi F, Guzzetti S, et al. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. Circ Res 1986 Aug;59(2):178-193. [doi: 10.1161/01.res.59.2.178] [Medline: 2874900]
26. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Front Physiol 2013;4:26. [doi: 10.3389/fphys.2013.00026] [Medline: 23431279]
27. Eckberg DL. Sympathovagal balance: a critical appraisal. Circulation 1997 Nov 4;96(9):3224-3232. [doi: 10.1161/01.cir.96.9.3224] [Medline: 9386196]
28. Berger RD, Saul JP, Cohen RJ. Transfer function analysis of autonomic regulation. I. Canine atrial rate response. Am J Physiol 1989 Jan;256(1 Pt 2):H142-H152. [doi: 10.1152/ajpheart.1989.256.1.H142] [Medline: 2912176]
29. Porta A, Aletti F, Vallais F, Baselli G. Multimodal signal processing for the analysis of cardiovascular variability. Philos Trans A Math Phys Eng Sci 2009 Jan 28;367(1887):391-409. [doi: 10.1098/rsta.2008.0229] [Medline: 18940775]
30. Jo JA, Blasi A, Valladares E, Juarez R, Baydur A, Khoo MCK. Model-based assessment of autonomic control in obstructive sleep apnea syndrome during sleep. Am J Respir Crit Care Med 2003 Jan 15;167(2):128-136. [doi: 10.1164/rccm.200202-096OC] [Medline: 12406844]
31. Zhang J, Mani I. KNN approach to unbalanced data distributions: a case study involving information extraction. Presented at: Proceedings of the ICML'2003 - Workshop on Learning from Imbalanced Datasets; Aug 21, 2003 URL: https://www.bibsonomy.org/bibtex/2cf4d2ac8bdac874b3d4841b4645a5a90/diana [accessed 2026-02-03]
32. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-357. [doi: 10.1613/jair.953]
33. Graf ABA, Borer S. Normalization in support vector machines. Presented at: Pattern Recognition: 23rd DAGM Symposium; Sep 10-12, 2001. [doi: 10.1007/3-540-45404-7_37]
34. Jurafsky D, Martin JH. Logistic regression. In: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition 2026. URL: https://web.stanford.edu/~jurafsky/slp3 [accessed 2026-02-10]
35. Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2020 Dec;97(Part B):105524. [doi: 10.1016/j.asoc.2019.105524]
36. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013 Jan;36(1):27-46. [doi: 10.1111/j.1600-0587.2012.07348.x]
37. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157-1182 [FREE Full text]
38. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Presented at: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95); Aug 20-25, 1995; Montréal, QC p. 1137-1143 URL: https://ai.stanford.edu/~ronnyk/accEst.pdf [accessed 2026-02-03]

XSL•FO
RenderX

39. Marmarelis VZ, Shin DC, Kang Y, Novak V. Data-based modeling of cerebral hemodynamics quantifies impairment of cerebral blood flow regulation in type-2 diabetes. J Cereb Blood Flow Metab 2024 Nov;44(11):1288-1301. [doi: 10.1177/0271678X241254716] [Medline: 38748923]
40. Da Silva CD, Catai AM, Abreu RD, et al. Cardiorespiratory coupling as an early marker of cardiac autonomic dysfunction in type 2 diabetes mellitus patients. Respir Physiol Neurobiol 2023 May;311:104042. [doi: 10.1016/j.resp.2023.104042] [Medline: 36858335]
41. Bari V, Vaini E, Pistuddi V, et al. Comparison of causal and non-causal strategies for the assessment of baroreflex sensitivity in predicting acute kidney dysfunction after coronary artery bypass grafting. Front Physiol 2019;10:1319. [doi: 10.3389/fphys.2019.01319] [Medline: 31681021]
42. Nayan NM, Islam A, Islam MU, Ahmed E, Hossain MM, Alam MZ. SMOTE oversampling and Near Miss undersampling based diabetes diagnosis from imbalanced dataset with XAI visualization. Presented at: 2023 IEEE Symposium on Computers and Communications (ISCC); Jul 9-12, 2023. [doi: 10.1109/ISCC58397.2023.10218281]
43. Pop-Busui R, Boulton AJM, Feldman EL, et al. Diabetic neuropathy: a position statement by the American Diabetes Association. Diabetes Care 2017 Jan;40(1):136-154. [doi: 10.2337/dc16-2042] [Medline: 27999003]
44. Rathod S, Phadke L, Chaskar U, Patil C. Heart rate variability measured during rest and after orthostatic challenge to detect autonomic dysfunction in type 2 diabetes mellitus using the classification and regression tree model. Technol Health Care 2022;30(2):361-378. [doi: 10.3233/THC-213048] [Medline: 34250917]
45. G S, Kp S, R V. Automated detection of diabetes using CNN and CNN–LSTM network and heart rate signals. Procedia Comput Sci 2018;132:1253-1262. [doi: 10.1016/j.procs.2018.05.041]
46. Swapna G, Vinayakumar R, Soman KP. Diabetes detection using deep learning algorithms. ICT Express 2018 Dec;4(4):243-246. [doi: 10.1016/j.icte.2018.10.005]
47. Research. Flavia M G S Oliveira. URL: https://flavia.ene.unb.br/research [accessed 2026-02-17]

Abbreviations

AUC-ROC: area under the receiver operating characteristic curve
BIDMC: Beth Israel Deaconess Medical Center
CAN: cardiovascular autonomic neuropathy
CARTs: cardiovascular autonomic reflex tests
CBV: cerebral blood velocity
CNN–LSTM: convolutional neural network–long short-term memory
ECG: electrocardiogram
FRF: frequency response function
HF: high-frequency
HRV: heart rate variability
ILV: instantaneous lung volume
IR: impulse response
IRB: institutional review board
LF: low-frequency
LR: logistic regression
ML: machine learning
NM: NearMiss-1
RBF: radial basis function
RCC: respiratory-cardiac coupling
RRI: R-to-R interval
SMOTE: Synthetic Minority Oversampling Technique
SVM: support vector machine
T2DM: type 2 diabetes mellitus

Edited by S Li; submitted 08.Aug.2025; peer-reviewed by CA Sanches, H Selvitopi, S Eshun; revised version received 09.Dec.2025; accepted 29.Dec.2025; published 23.Feb.2026.

Please cite as:
Oliveira FMGSA, Cavalcanti SM, Khoo MCK
Cardiorespiratory Markers of Type 2 Diabetes: Machine Learning–Based Analysis
JMIR Diabetes 2026;11:e82084
URL: https://diabetes.jmir.org/2026/1/e82084
doi:10.2196/82084

XSL•FO
RenderX

© Flavia Maria G S A Oliveira, Sandro Muniz Cavalcanti, Michael C K Khoo. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predictors of Glycemic Response to Sulfonylurea Therapy in Type 2 Diabetes Over 12 Months: Comparative Analysis of Linear Regression and Machine Learning Models

Shilpa Garg[1], BTech; Robert Kitchen[2], PhD; Ramneek Gupta[2,3], PhD; Emanuele Trucco[4], PhD; Ewan Pearson[1], MA, MB, BChir, PhD

[1]Diabetes Endocrinology and Reproductive Biology, School of Medicine, University of Dundee, Nethergate, Dundee, Dundee, United Kingdom

[2]Novo Nordisk, Oxford, United Kingdom

[3]Disease Intelligence Pte Ltd, 10 Anson Road, Singapore, Singapore

[4]Computing Department, School of Science and Engineering, University of Dundee, Dundee, United Kingdom

Corresponding Author:
Shilpa Garg, BTech
Diabetes Endocrinology and Reproductive Biology, School of Medicine, University of Dundee, Nethergate, Dundee, Dundee, United Kingdom

Abstract

Background: Sulfonylureas are commonly prescribed for managing type 2 diabetes, yet treatment responses vary significantly among individuals. Although advances in machine learning (ML) may enhance predictive capabilities compared to traditional statistical methods, their practical utility in real-world clinical environments remains uncertain.

Objective: This study aimed to evaluate and compare the predictive performance of linear regression models with several ML approaches for predicting glycemic response to sulfonylurea therapy using routine clinical data, and to assess model interpretability using Shapley Additive Explanations (SHAP) analysis as a secondary analysis.

Methods: A cohort of 7557 individuals with type 2 diabetes who initiated sulfonylurea therapy was analyzed, with all patients followed for 1 year. Linear and logistic regression models were used as baseline comparisons. A range of ML models was trained to predict the continuous change in hemoglobin A_{1c} (HbA_{1c}) levels and the achievement of HbA_{1c} <58 mmol/mol at follow-up. These models included random forest, extreme gradient boosting, support vector machines, a conventional feedforward neural network, and Bayesian additive regression trees. Model performance was assessed using standard metrics including R^2 and root mean squared error for regression tasks and area under the receiver operating characteristic for classification. In a subset of 2361 patients, nonfasting connecting peptide (C-peptide) was analyzed as a proxy for β-cell function. SHAP analysis was performed to identify and compare key predictors driving model performance across methods.

Results: All models exhibited similar performance, with no significant advantages of ML techniques over linear regression. For continuous outcomes, Bayesian additive regression trees demonstrated the highest R^2 (0.445) and lowest root mean squared error (0.105), though the differences among models were minimal. For the binary outcome, extreme gradient boosting achieved the highest area under the receiver operating characteristic curve (0.712), with CIs overlapping those of other models. Across all models, baseline HbA_{1c} was consistently the primary predictor, explaining the majority of the variance. SHAP analyses confirmed that baseline HbA_{1c}, age, BMI, and sex were the most influential predictors. Sensitivity analyses and hyperparameter tuning did not significantly improve model performance. In the C-peptide subset, higher C-peptide levels were associated with greater glycemic improvement (β=−3.2 mmol/mol per log(C-peptide); P<.001).

Conclusions: In this large, population-based cohort, ML models did not outperform traditional regression for predicting glycemic response to sulfonylureas. These findings suggest that limited gains from ML likely reflect an absence of strong nonlinear or high-order interactions in routine clinical data and that available features may not capture sufficient biological heterogeneity for complex models to confer added benefit. The inclusion of a C-peptide subset provides additional mechanistic insight by linking preserved β-cell function with treatment response.

(JMIR Diabetes 2026;11:e82635) doi:10.2196/82635

KEYWORDS

drug response; glycated hemoglobin; linear regression; machine learning models; treatment response prediction; type 2 diabetes

XSL•FO
RenderX

Introduction

Sulfonylureas are among the most commonly prescribed classes of glucose-lowering medications for individuals with type 2 diabetes. Their cost-effectiveness and accessibility make them particularly valuable in resource-constrained settings [1]. However, significant variability exists in glycemic responses among individuals. This variability is influenced by various clinical and biological factors, such as age, kidney function, and genetic predispositions [2,3]. Identifying predictors of treatment response is essential for advancing precision medicine approaches and minimizing trial-and-error prescribing practices [4].

Because sulfonylureas lower glucose primarily by stimulating insulin secretion from pancreatic β-cells, the degree of preserved β-cell function, often estimated by circulating connecting peptide (C-peptide) [5], may influence treatment response. However, such markers are rarely available in real-world datasets and are not routinely included in prediction studies.

Machine learning (ML) methods have shown promise in predicting treatment responses more accurately than traditional regression models, particularly due to their ability to handle complex, nonlinear interactions between variables without requiring prespecified assumptions [6,7]. In this context, ML approaches can capture subtle, multidimensional relationships that may be overlooked by traditional models, efficiently process large-scale longitudinal data, and generate data-driven insights that inform treatment selection. ML also offers better integration of diverse data types and improved interpretability through explainable AI, increasing clinical applicability [8]. Despite this promise, relatively few studies have focused on modeling glycemic response in diabetes using real-world data [9]. This gap in research presents a significant opportunity for further investigation.

Here, we use sulfonylurea response as an exemplar of diabetes drug response, due to its widespread use and the availability of clinical data. We evaluate and compare the efficacy of 5 ML models, including random forest, support vector machines, extreme gradient boosting (XGBoost), a conventional feedforward neural network (NN), and Bayesian additive regression trees (BART), in predicting the glycemic response to sulfonylureas in patients with type 2 diabetes. These models are compared with standard linear and logistic regression for continuous (change in hemoglobin A_{1c} [HbA_{1c}]) and binary (achievement of HbA_{1c} <58 mmol/mol) outcomes. Analyses were conducted using a large real-world cohort from the GoDARTS (Genetics of Diabetes Audit and Research in Tayside Scotland) study, including a biologically informative subset with C-peptide measurements to assess the contribution of β-cell function.

In addition to comparing predictive performance across models, we conducted a secondary analysis to examine feature contributions using Shapley Additive Explanations (SHAP) [10]. This analysis allowed us to determine whether ML-derived feature importance aligns with the predictors identified by traditional regression approaches, providing insight into clinical interpretability and the practical utility of ML for informing treatment choice.

Methods

Study Population

The data were obtained from the GoDARTS [11]. This population-based cohort links prescription, clinical, and laboratory records for individuals with diabetes in Tayside and Fife. The inclusion criteria included patients with type 2 diabetes who initiated sulfonylurea therapy (either as monotherapy or in combination), had a baseline HbA_{1c} measurement (defined as the closest value within 183 days before to 7 days after treatment initiation), and a follow-up HbA_{1c} measurement after a 1-year period. The 183-day window was selected to balance data availability and clinical relevance. For this analysis, only 2 HbA_{1c} values per patient were used, 1 at baseline and 1 at follow-up, in line with the model's aim of predicting glycemic response from initial clinical features.

Ethical Considerations

Data provision and linkage were carried out by the University of Dundee Health Informatics Centre, with analysis of anonymized data performed in an ISO27001 and Scottish Government–accredited secure safe haven. Health Informatics Centre standard operating procedures were reviewed and approved by the National Health Service (NHS) East of Scotland Research Ethics Service (22/ES/0034), and consent for this study was obtained from the NHS Fife Caldicott Guardian. Under these approvals, secondary analysis of anonymised routine healthcare data does not require additional participant consent or compensation.

Baseline Predictor Variables

Baseline clinical features included age, sex, HbA_{1c}, BMI, total cholesterol, high-density lipoprotein (HDL) cholesterol, smoking status, systolic blood pressure, alkaline phosphatase, alanine transaminase, serum potassium, serum creatinine, bilirubin, and albumin. These variables were selected based on their availability in routine care and their known or suspected relevance to glycemic outcomes [12]. Except for baseline HbA_{1c} (as defined above), all measurements were defined as the closest recorded value within 2 years before to 90 days after sulfonylurea initiation.

To estimate average daily sulfonylurea dose, prescription records were used to extract drug strength and quantity dispensed. Five sulfonylureas were included: gliclazide, glipizide, glimepiride, glibenclamide, and tolbutamide. Each prescription's dose was standardized by dividing the prescribed dose by the drug's maximum recommended daily dose (as per the British National Formulary). This yielded a standardized dose unit, which was then multiplied by the number of tablets prescribed per prescription to calculate the total standardized dose. For each patient, the total dose was summed across all prescriptions, excluding the last one, and divided by treatment duration to derive the average daily dose. This dose was then categorized into low, medium, and high using quartiles.

XSL•FO
RenderX

Outcome Definitions

The primary continuous outcome was defined as the change in glycated hemoglobin (HbA_{1c}), measured in millimoles per mole, from baseline (at the time of sulfonylurea initiation) to the follow-up measurement closest to 12 months, within a window of 6-15 months.

The binary outcome was defined as whether a patient achieved a follow-up HbA_{1c} level below 58 mmol/mol.

Data Preparation

To ensure consistency and compatibility with ML models, several preprocessing steps were applied. Continuous variables with skewed distributions underwent log transformation to approximate a normal distribution [13], enhancing model stability and reducing the influence of extreme values. Following this, all continuous predictors, including laboratory test results and physiological measurements, were scaled to a range between 0 and 1 using min-max normalization [14]. This rescaling placed variables on a uniform scale, which is particularly important for algorithms like NNs that are sensitive to variable magnitudes. Categorical variables (eg, sex, smoking status, treatment group, average daily dose) were converted using one-hot encoding to make them compatible with model inputs.

Missing Data Imputation and Collinearity Assessment

Patients missing either baseline or follow-up HbA_{1c} measurements were excluded. For remaining clinical predictors, missingness was below 10% and not clustered within specific individuals. Missing values were imputed using multiple imputation by chained equations [15] implemented in R (mice v3.18.0). Five imputed datasets were generated with 50 iterations each, using predictive mean matching for continuous variables. Full details of the imputation model are provided in Section 1 in Multimedia Appendix 1. Convergence was assessed using the mean and variance of each variable across iterations and comparing distributions of observed and imputed values. Analyses were performed on pooled estimates derived using Rubin's rules [16].

Collinearity among predictors was evaluated using variance inflation factors (VIFs) [17]. Predictors with VIF values greater than 5 were reviewed for redundancy. In our final models, VIFs ranged from 1.06 to 1.5, indicating no meaningful multicollinearity. As a sensitivity check, strongly correlated clinical variables ($r > 0.8$) were examined, and when overlap occurred (eg, estimated glomerular filtration rate vs serum creatinine), the variable more routinely and reliably measured in clinical practice (serum creatinine) was retained.

Statistical Analysis: Baseline Models

Initial statistical analyses were conducted using linear regression [18] for the continuous outcome and logistic regression for the binary outcome. These models identified baseline associations between clinical predictors and glycemic response to sulfonylurea therapy. Logistic regression estimated the probability of achieving an HbA_{1c} <58 mmol/mol. Of the 7557 individuals included, 3818 achieved the target, and 3739 did not.

Residualization of Baseline HbA_{1c}

To disentangle treatment response from baseline glycemia, change in HbA_{1c} was regressed on baseline HbA_{1c}. The residuals from this model were used as outcomes for ML analyses. This allowed the identification of predictors influencing glycemic response independent of baseline HbA_{1c} levels [19].

ML Models

Five ML models were implemented, reflecting diverse algorithmic strategies:

1. Random forest: An ensemble method that constructs multiple decision trees on bootstrapped data and aggregates their predictions [20].
2. Support vector machines: A kernel-based classifier that establishes an optimal separating hyperplane [21].
3. XGBoost: A boosting technique that sequentially builds trees to minimize residual error [22].
4. NNs: A conventional feedforward NN (multilayer perceptron) trained using resilient backpropagation. Comprising layers of interconnected neurons, NNs excel at modeling complex relationships and require larger datasets and regularization to mitigate overfitting [23].
5. BARTs: A Bayesian ensemble method that combines multiple regression trees and estimates uncertainty in predictions [24]. BART is noted for strong performance in clinical applications [25-27].

Model Implementation

For model development, a 2-stage validation framework combining cross-validation and a held-out test set was used. The data were randomly split into a 70% training set and a 30% held-out test set. Within the training set, 10-fold cross-validation [18] was used for hyperparameter tuning and model selection to enhance model stability and reduce overfitting. Final performance was evaluated on the held-out test set, which remained unseen during training.

All analyses were performed in R (version 4.3.0). A detailed description of the packages and functions used for each model is presented in Section 2 in Multimedia Appendix 1, and XGBoost and NN hyperparameters are provided in Sections 3 and 4 in Multimedia Appendix 1, respectively. All preprocessing and modeling code is publicly available on GitHub [28].

Feature Importance

To identify the clinical features most strongly influencing model predictions, we assessed feature importance using SHAP values along with the built-in variable importance metrics from each model. SHAP values quantify the contribution of individual predictors to model outputs, enabling transparent, model-agnostic interpretation. Although SHAP can be applied to multiple model types, our results focused on the XGBoost model because it showed optimal predictive performance. SHAP summary plots were generated to visualize the magnitude and direction of feature effects, ranking predictors by their mean absolute SHAP values. Comparative plots across models were generated to visualize predictor impact on treatment response. This unified approach supported consistent evaluation of feature relevance across models and enhanced clinical interpretability.

XSL•FO
RenderX

Performance Evaluation

Model performance was assessed separately for the continuous and binary outcomes. For the continuous outcome, evaluation metrics included root mean squared error (RMSE), mean absolute error, and the coefficient of determination (R^2), which indicates the proportion of variance in the outcome explained by the model [29,30].

For the binary outcome, performance was evaluated using standard classification metrics: area under the receiver operating characteristic curve (AUC), accuracy, sensitivity (recall), and specificity [31,32]. In the linear regression models, regression coefficients were interpreted to assess the direction and magnitude of each predictor's association with the outcome, while *P* values indicated the statistical significance of these associations. An R^2 value provided an overall measure of model fit, and a *P* value below .05 was considered statistically significant.

To evaluate differences in performance across models, a resampling-based approach was used to compare their predictive metrics [33]. Pairwise comparisons were conducted to assess whether any model significantly outperformed the others.

Results

Cohort Characteristics

The study included 7557 individuals with type 2 diabetes who initiated sulfonylurea therapy and had both baseline and follow-up HbA_{1c} values available. The cohort had a mean age of 63.7 (SD 11.8) years, and 57.9% (n=4377) were male. The mean baseline HbA_{1c} was 76.5 (SD 16.7) mmol/mol (Table 1).

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of the study population.

Clinical variable	Sulfonylurea cohort (N=7557)
Age at therapy initiation (y), mean (SD)	63.7 (11.8)
Sex, n (%)	
Male	4377 (57.9)
Female	3180 (42.1)
Average daily dose, n (%)	
Low	1844 (24.4)
Medium	3822 (50.6)
High	1891 (25)
Duration of diabetes (y), mean (SD)	4.96 (4.49)
Duration of diabetes, n (%)	
0 - 1 years	1416 (18.7)
1 - 5 years	3099 (41)
>5 years	3042 (40.3)
Time of treatment (mo), mean (SD)	11.4 (2.2)
Time from baseline HbA_{1c}[a] measurement to treatment start (d), mean (SD)	21.3 (29.1)
Year of drug start, mean (SD)	2010 (6.12)
BMI (kg/m^2), mean (SD)	31.3 (6.3)
Total cholesterol (mmol/L), mean (SD)	4.5 (1.2)
HDL[b] cholesterol (mmol/L), mean (SD)	1.2 (0.3)
Serum creatinine (µmol/L), mean (SD)	80.3 (27.4)
Albumin (g/L), mean (SD)	42.2 (4.0)
Bilirubin (µmol/L), mean (SD)	9.9 (5.2)
Alkaline phosphatase (U/L), mean (SD)	89.5 (42.1)
ALT/SGPT[c] (U/L), mean (SD)	34.2 (24.4)
Potassium (mmol/L), mean (SD)	4.4 (0.4)
Systolic blood pressure, mean (SD)	137 (17.4)
Smoking status, n (%)	
Ever smoked—yes	5628 (74.5)
Ever smoked—no	1870 (24.7)
Ever smoked—unknown	59 (0.8)
Therapy group, n (%)	
Mono	2508 (33.2)
Dual	4251 (56.3)
Triple	798 (10.6)
Index of multiple deprivation quintile, n (%)	
1 (most deprived)	1583 (16.5)
2	1609 (21.3)
3	1497 (19.8)
4	1396 (18.5)
5 (least deprived)	1246 (16.5)

XSL•FO
RenderX

Clinical variable	Sulfonylurea cohort (N=7557)
Unknown	226 (3.0)
Ethnicity, n (%)	
White	5696 (75.4)
Others/mixed	259 (3.4)
Missing	1602 (21.2)
Region, n (%)	
Tayside	5965 (78.9)
Fife	1592 (21.1)
Baseline HbA_{1c} (mmol/mol), mean (SD)	76.5 (16.7)
HbA_{1c} outcome (mmol/mol) (treatment HbA_{1c}), mean (SD)	61.1 (15.4)
HbA_{1c} response (change from baseline; mmol/mol), mean (SD)	–15.4 (18)

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]HDL: high-density lipoprotein.

[c]ALT/SGPT: alanine aminotransferase/serum glutamate pyruvate transaminase.

Associations Between Clinical Covariates and Treatment Response

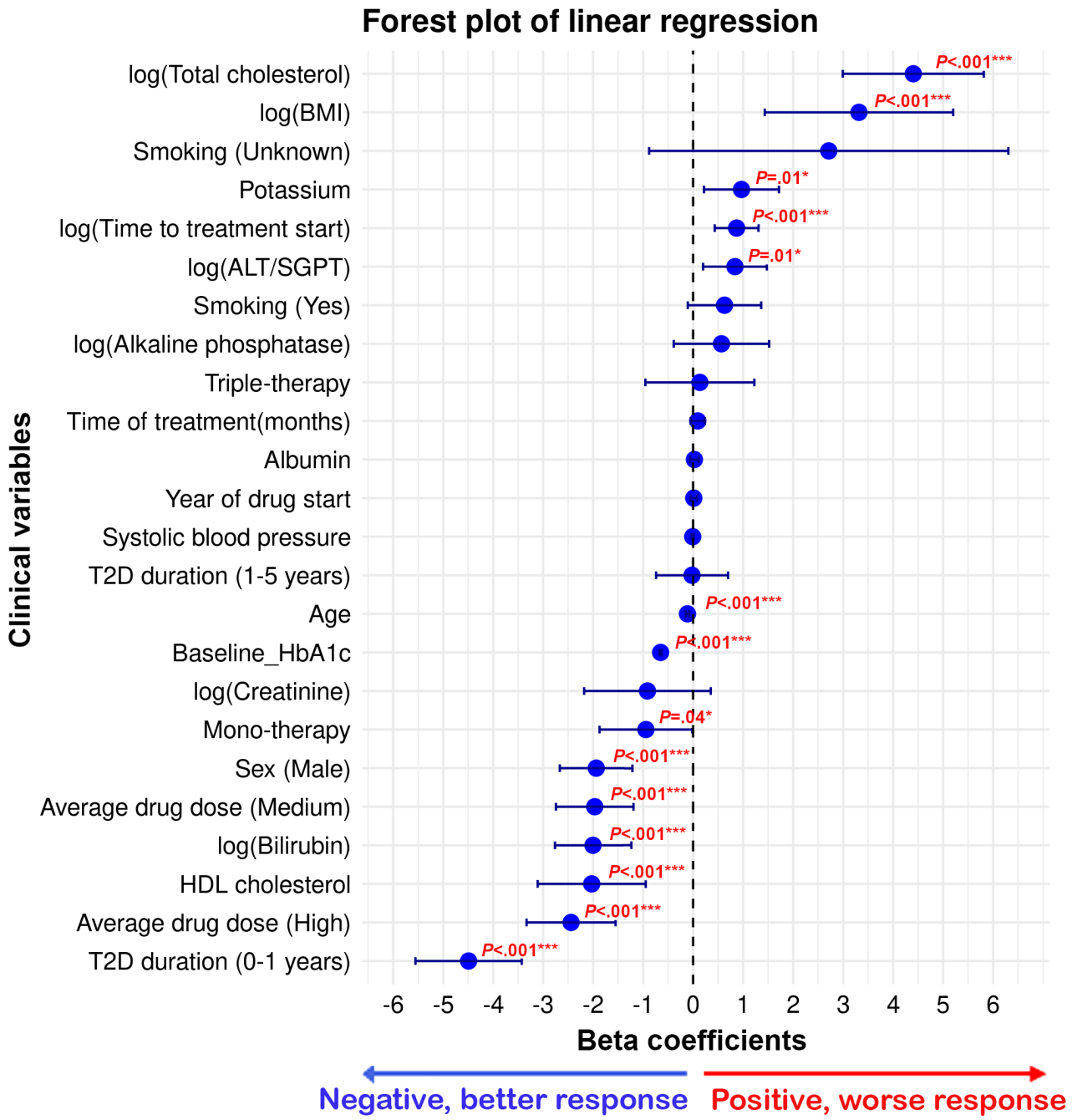
A linear regression model was fit to assess the relationship between clinical variables and the change in HbA_{1c}, defined as the difference between baseline HbA_{1c} and follow-up values (ie, change=treatment–baseline HbA_{1c}). A negative change in HbA_{1c} indicates a better treatment response to sulfonylureas, while a positive change signifies a worse response. No feature scaling (min-max normalization) was applied to the variables in this model. The model demonstrated satisfactory fit, with an R^2 value of 0.41 and a significant F-statistic (222.2; $P<.001$).

The linear regression analysis identified several clinical variables significantly associated with the change in HbA_{1c} among individuals treated with sulfonylureas. Older age and higher baseline HbA_{1c} were associated with greater HbA_{1c} reductions (negative coefficients), indicating more favorable responses. In contrast, a higher BMI was associated with smaller reductions, suggesting that individuals with higher BMI may struggle to achieve desired glycemic control. Additionally, sex differences indicated that male participants demonstrated slightly better response to sulfonylurea treatment than female participants.

These associations are illustrated in Figure 1, which presents a forest plot of the linear regression coefficients and CIs, highlighting the magnitude and direction of each predictor's effect on HbA_{1c} change.

XSL•FO
RenderX

Figure 1. Forest plot showing regression coefficients and 95% CIs for predictors of change in hemoglobin A_{1c} (HbA_{1c}). Points represent model estimates and horizontal lines indicate 95% CIs. ALT: alanine aminotransferase; SGPT: alanine aminotransferase/serum glutamate pyruvate transaminase; T2D: type 2 diabetes. *P<.05; **P<.01; ***P<.001.



C-Peptide Analysis

In a subset of 2361 individuals with nonfasting C-peptide data, higher C-peptide levels were strongly associated with greater reductions in HbA_{1c} at 12 months (linear regression: $\beta=-3.2$ mmol/mol per log(C-peptide); P<.001). This finding suggests that preserved endogenous insulin secretion contributes to more favorable treatment outcomes.

ML Model Performance for Continuous Outcome

For the continuous outcome of predicting changes in HbA_{1c}, several models were evaluated. The results indicate that the BART model exhibited the lowest RMSE of 0.105 (21%) and the lowest mean absolute error of 0.079 (16.1%), highlighting its effective performance in estimating continuous changes. XGBoost and NNs also performed comparably, with RMSE values of 0.106.

On the original HbA_{1c} scale, this corresponds to an approximate prediction error of 13.8 mmol/mol, comparable to the residual standard error from the linear regression model. Thus, the clinical prediction error was approximately ~14 mmol/mol.

However, the differences in RMSE and R^2 across all models were minimal, indicating that while BART performed slightly better, the performance of all models was relatively comparable

XSL•FO
RenderX

(Table 2). The similar R^2 and RMSE values further suggest that no single model stands out significantly.

Table . Regression model performance metrics (root mean squared error [RMSE], mean absolute error [MAE], and R^2) for continuous outcome prediction across all 6 models[a].

Models	RMSE	MAE	R^2
Linear regression	0.106	0.08	0.434
RF[b]	0.108	0.082	0.424
SVM[c]	0.106	0.079	0.438
XGBoost[d]	0.106	0.08	0.433
NN[e]	0.106	0.081	0.427
BART[f]	0.105	0.079	0.445

[a]RMSE is shown as the normalized values.
[b]RF: random forest.
[c]SVM: support vector machine.
[d]XGBoost: extreme gradient boosting.
[e]NN: neural network.
[f]BART: Bayesian additive regression trees.

Statistical Comparison of Model Performance

In addition to reporting standard performance metrics, statistical comparisons were performed using resampling-based techniques. Pairwise comparisons of RMSE and R^2 values across all models showed no statistically significant differences in performance; all ML models, including the linear regression baseline, performed similarly on this dataset.

Sensitivity Analysis: Residualized HbA_{1c} Change

A sensitivity analysis was performed to evaluate predictors of HbA_{1c} change independent of baseline glycemia. Across all models, the maximum R^2 value decreased to 0.05, indicating that only ~5% of the residual variance in 12-month HbA_{1c} response was explained by routine clinical features after removing the effect of baseline HbA_{1c}.

The performance metrics from this analysis further indicated that the RMSE and R^2 values remained consistent across most models (Table 3). However, XGBoost and BART showed poorer performance, with high RMSE and lower R^2 values. This likely reflects the fact that these algorithms are better suited for large, high-dimensional, or highly nonlinear datasets, whereas the present dataset may not contain sufficient complexity. This consistency across most models suggests that while some approaches explain marginally more variance in the sensitivity analysis, their predictive accuracy in terms of mean squared error remains stable.

Table . Model performance after adjustment for baseline hemoglobin A_{1c} (HbA_{1c}) across all 6 models.

Models	RMSE[a]	MAE[b]	R^2
Linear regression	0.127	0.095	0.054
RF[c]	0.126	0.095	0.056
SVM[d]	0.128	0.094	0.051
XGBoost[e]	0.230	0.183	0.01
NN[f]	0.127	0.095	0.053
BART[g]	0.214	0.191	0.021

[a]RMSE: root mean squared error.
[b]MAE: mean absolute error.
[c]RF: random forest.
[d]SVM: support vector machine.
[e]XGBoost: extreme gradient boosting.
[f]NN: neural network.
[g]BART: Bayesian additive regression trees.

XSL•FO
RenderX

ML Model Performance for Binary Outcome

For the binary outcome of predicting achievement of HbA_{1c} <58 mmol/mol, model performance was assessed using the AUC and accuracy. The XGBoost model achieved the highest AUC (0.712), followed closely by BART (0.710), indicating modest discriminatory ability. Logistic regression performed similarly, with an AUC of 0.702.

The CIs for all models showed substantial overlap (ranging from 0.681 to 0.724 for logistic regression and 0.692 to 0.733 for XGBoost), indicating that no model demonstrated statistically superior discrimination. Overall, the models were broadly comparable in their ability to distinguish responders from nonresponders.

Model-level classification metrics are summarized in Table 4, and the corresponding ROC curves for all models are shown in Figure 2, illustrating their similar performance profiles. In Figure 2, colored curves represent the individual models, visually reinforcing the overlapping AUCs and the absence of meaningful differences in classification performance.

Table . Discrimination and classification performance of binary outcome models.

Models	AUC[a] (95% CI)	Accuracy	Precision	Recall
Logistic regression	0.702 (0.681 - 0.724)	0.654	0.657	0.628
RF[b]	0.708 (0.687 - 0.729)	0.652	0.656	0.628
SVM[c]	0.705 (0.684 - 0.727)	0.65	0.656	0.618
XGBoost[d]	0.712 (0.692 - 0.733)	0.646	0.65	0.625
NN[e]	0.699 (0.678 - 0.72)	0.645	0.645	0.636
BART[f]	0.71 (0.689 - 0.731)	0.651	0.652	636

[a]AUC: area under the receiver operating characteristic curve.

[b]RF: random forest.

[c]SVM: support vector machine.

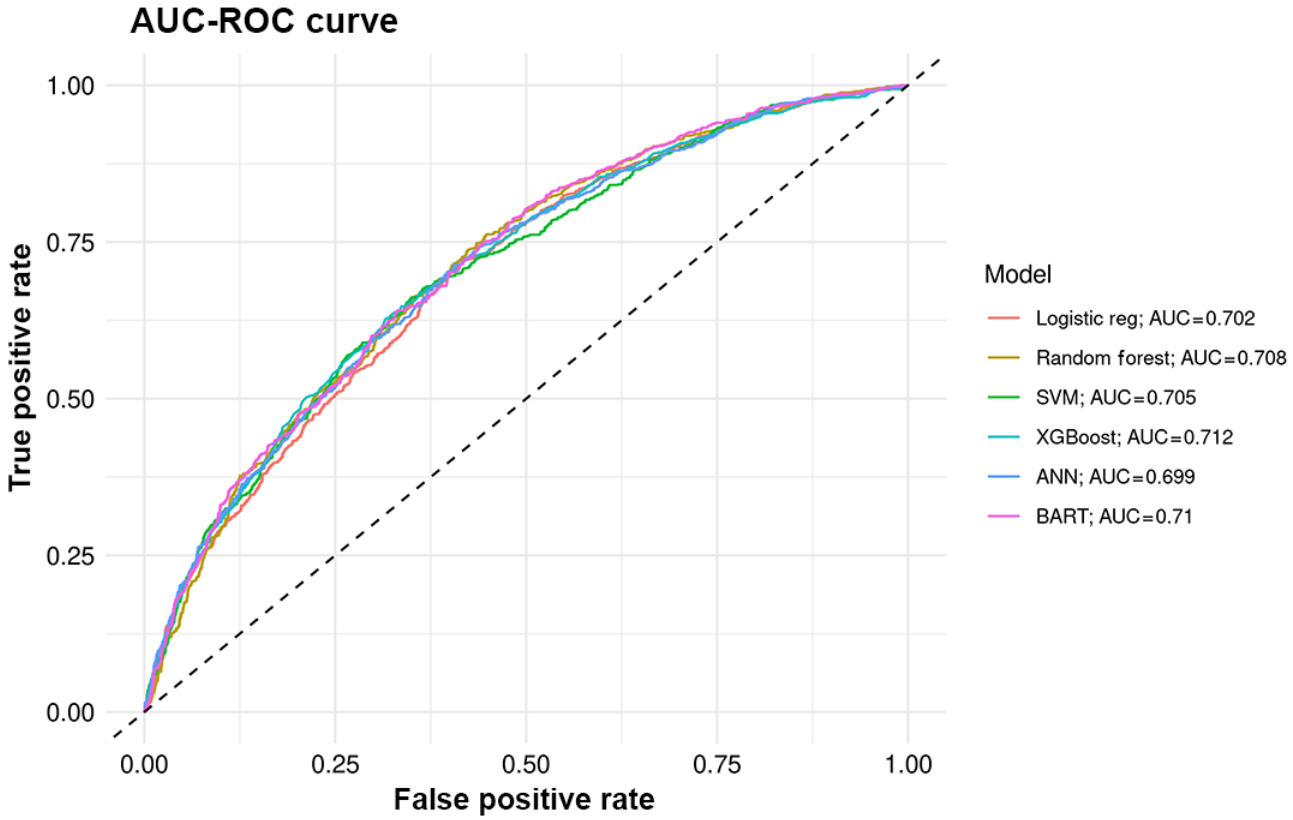
[d]XGBoost: extreme gradient boosting.

[e]NN: neural network.

[f]BART: Bayesian additive regression trees.

XSL•FO
RenderX

Figure 2. Receiver operating characteristic (ROC) curves for binary outcome prediction models. Colors correspond to individual models as shown in the legend. ANN: artificial neural network; AUC: area under the receiver operating characteristic curve; BART: Bayesian additive regression trees; SVM: support vector machine; XGBoost: extreme gradient boosting.



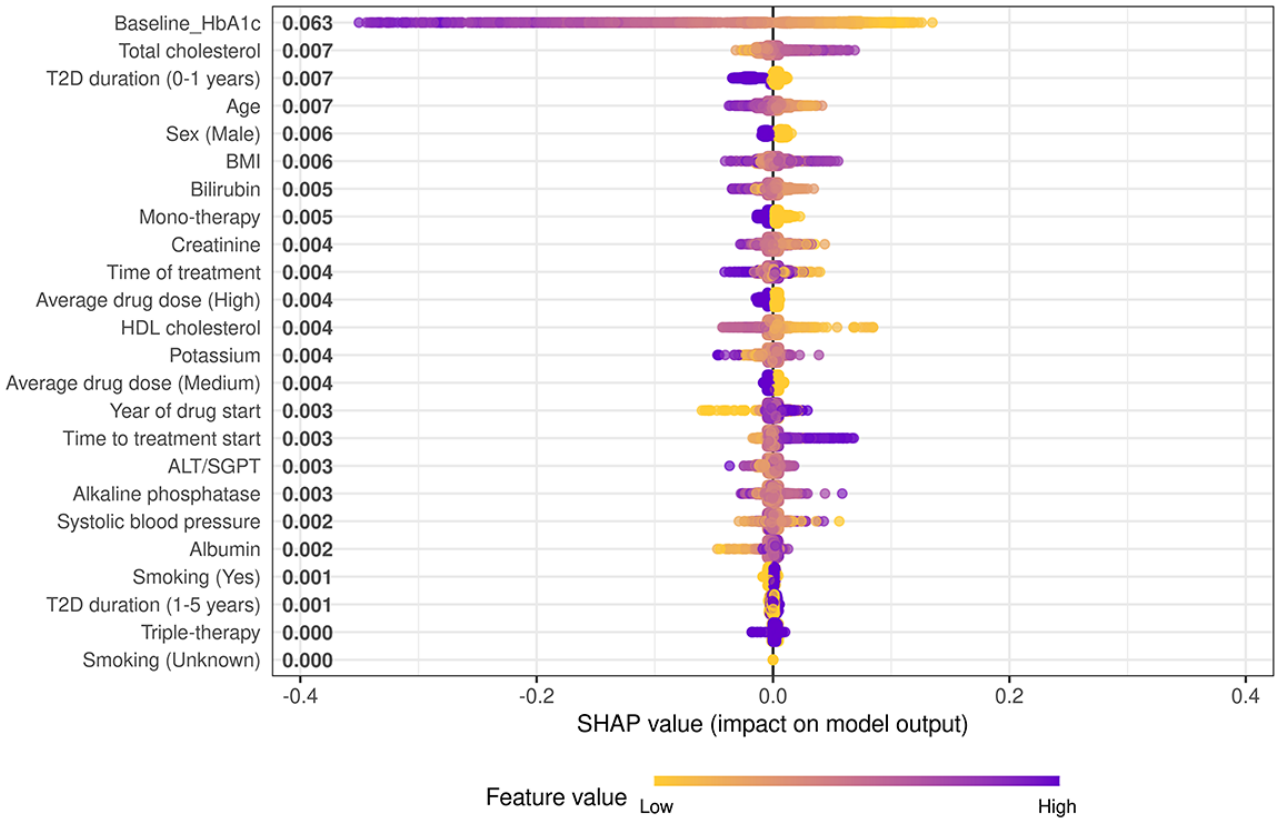
Feature Importance and SHAP Interpretability

Feature importance analyses consistently identified baseline HbA_{1c} as the most significant predictor across all models. Other variables such as BMI, alanine transaminase, total cholesterol, and systolic blood pressure were also found to be significant, though their rankings varied slightly between algorithms. Across all 5 models, the rankings of key predictors remained largely consistent.

To further explain feature contributions, a SHAP summary plot derived from the XGBoost model is presented (Figure 3), offering a more granular view of individual feature effects on model predictions. Baseline HbA_{1c} had the highest mean SHAP value (0.063), followed by total cholesterol, duration of diabetes, and age. Higher baseline HbA_{1c} values were associated with larger predicted reductions in HbA_{1c} (ie, more negative SHAP values), indicating greater expected treatment benefit.

XSL•FO
RenderX

Figure 3. Shapley Additive Explanations (SHAP) summary plot of feature importance in predicting glycemic response. Each dot represents 1 patient. The x-axis indicates the SHAP value (impact on model output). The color gradient reflects feature values (blue=higher values; yellow=lower values). Features are ordered by mean absolute SHAP values, indicating overall contribution to model predictions. ALT/SGPT: alanine aminotransferase/serum glutamate pyruvate transaminase; HbA_{1c}: hemoglobin A_{1c}; HDL: high-density lipoprotein; T2D: type 2 diabetes.



The SHAP plot further shows that lower total cholesterol values corresponded to more negative SHAP values, suggesting better predicted outcomes, whereas higher cholesterol was linked to reduced response. Similarly, shorter diabetes duration and younger age were associated with more favorable predictions. For the variable sex, blue points represent male participants and yellow points represent female participants; male participants were associated with more negative SHAP values, indicating a better predicted response, compared to female participants, whose SHAP values clustered closer to or above zero.

Discussion

Principal Findings

This study compared the predictive performance of traditional regression models and a range of ML algorithms for predicting glycemic response to sulfonylurea therapy in individuals with type 2 diabetes. The primary finding is that, with the dataset used, all models demonstrated comparable predictive performance. No ML approach significantly outperformed standard regression for either the continuous outcome or the binary outcome. These results indicate that, within routinely collected clinical data, the additional algorithmic complexity of ML methods does not necessarily yield superior predictive accuracy. Regression therefore remains a robust and interpretable option for predicting drug response in this context.

Linear regression analysis revealed that the model explained approximately 43% of the variance in changes to HbA_{1c}. In the sensitivity analysis, after adjusting for baseline HbA_{1c}, the maximum R^2 across all models dropped to 0.05, indicating that only a small proportion of outcome variability was captured by the remaining routine clinical features. This highlights the need for additional or more informative biomarkers to improve prediction.

Additionally, only about 50% (n=3818) of the participants achieved glycemic control after 1 year of sulfonylurea therapy, despite the relatively homogeneous clinical characteristics of the cohort. This finding highlights considerable interindividual variability in treatment response, suggesting that additional biological and behavioral factors may shape drug efficacy. Such heterogeneity may reflect differences in pharmacodynamic sensitivity, medication adherence, β-cell reserve, and underlying insulin resistance. BMI and HDL were considered indirect proxies of insulin resistance, as higher BMI is typically associated with greater insulin resistance, whereas higher HDL levels are generally linked to improved insulin sensitivity. Consistent with this, participants with higher BMI had poorer glycemic response, while those with higher HDL tended to show more favorable outcomes. However, direct measures of insulin resistance were not available in this dataset.

XSL•FO
RenderX

No Added Value From ML Methods

While multiple ML algorithms were evaluated in parallel with traditional regression models, none demonstrated superior predictive performance. Across both continuous and binary outcomes (Tables 2 and 4), differences in metrics such as RMSE, R^2, and AUC were small and not statistically significant, with overlapping CIs for all models. Even after hyperparameter tuning, predictive metrics remained modest, suggesting that ML methods did not uncover hidden patterns or interactions that traditional models missed.

This limited gain in predictive accuracy likely reflects the nature of routinely collected clinical data, which may lack sufficient biological complexity for ML algorithms to exploit. When input variables do not encompass detailed mechanistic information, even advanced algorithms cannot extract additional predictive signal. Consequently, transparent and easily interpretable models, such as linear or mixed-effects regression, may remain preferable, particularly when predictive performance is comparable. These models allow clinicians to understand feature contributions directly and translate findings into actionable treatment decisions.

Although complex ML models theoretically enable the capture of nonlinear relationships, their greater computational burden and reduced interpretability may limit their clinical utility unless they provide meaningful improvements in accuracy. The consistency of results across all modeling strategies, ranging from simple linear regression to ensemble and NN approaches, suggests that the available clinical features may not contain enough biological heterogeneity for ML methods to offer an advantage.

By intentionally comparing models of differing complexity, this study demonstrates that when data lack substantial nonlinearity or high-dimensional interactions, regression-based methods may remain more appropriate and efficient. This finding supports the continued reliance on interpretable models in routine clinical prediction tasks, where model parsimony and interpretability remain more valuable than algorithmic complexity for precision-medicine applications.

Features That Inform Drug Response Prediction

Feature importance and SHAP analyses consistently identified baseline HbA_{1c}, age, BMI, and sex as key predictors across regression and ML models. Baseline HbA_{1c} was the strongest predictor, reflecting both regression to the mean and true physiological responsiveness [34]. Older age and male sex were associated with greater HbA_{1c} reduction, whereas higher BMI predicted poorer response, aligning with evidence that adiposity may reduce sulfonylurea effectiveness [35]. These findings parallel results from the 5-drug predictive model developed by Dennis et al [36], suggesting that these core predictors generalize across therapeutic classes.

C-peptide, available for a subset of participants, showed a strong positive association with glycemic improvement, consistent with the insulin-secretagogue mechanism of sulfonylureas. This highlights the contribution of β-cell reserve to treatment heterogeneity and underscores the value of including mechanistic biomarkers to enhance model interpretability and predictive accuracy.

Limitations

This study has several limitations. First, routinely collected clinical data omit key determinants of treatment response such as adherence, diet, physical activity, genetics, and social factors. The limited availability of C-peptide prevented fuller assessment of β-cell function, and its strong association with response suggests that the inclusion of mechanistic biomarkers would likely improve predictive accuracy.

Second, the study population was limited to patients from Tayside and Fife in Scotland, which may reduce the generalizability of the findings to other regions or health care systems with different population characteristics or clinical practices.

Third, treatment response was assessed using a single HbA_{1c} value taken between 6 and 15 months after treatment initiation. Although data closest to 12 months post-initiation were used, variability in the follow-up period (6 - 15 mo) may introduce measurement variability and ought to be considered.

Additionally, the relatively low R^2 values across all models, even after applying rigorous methods such as a 70/30 train-test split and 10-fold cross-validation, suggest that the available clinical features alone do not explain sufficient variation in treatment response to support strong predictive performance.

Future Directions

Future research should focus on improving prediction models by incorporating richer and more diverse data sources, including genetic, metabolomic, and continuous glucose monitoring data, as well as direct measures of insulin resistance and β-cell function. Integrating these modalities could improve model accuracy and help explain why individuals respond differently to sulfonylurea treatment. In addition, future studies could explore advanced modeling approaches, such as deep learning or hybrid models that balance predictive power with ease of interpretation for clinical use. The increasing availability of real-world, longitudinal clinical data also supports the use of time-dependent models, such as recurrent neural networks or transformer models, to track how treatment response evolves over time. Finally, testing these models in independent and ethnically diverse populations will be important to assess their generalizability and real-world applicability.

In conclusion, this study shows that the traditional regression models remain robust, clinically interpretable, and sufficient for predicting glycemic response to sulfonylurea therapy using routine data. The comparable performance of ML methods suggests that model transparency and accessibility may currently outweigh the small gains offered by algorithmic complexity in this context.

XSL•FO
RenderX

Funding

This work was supported by funding from Novo Nordisk, which provided a PhD studentship to SG.

Authors' Contributions

EP, RK, and RG contributed to conceptualization and supervision. SG conducted the data analysis and wrote the original manuscript. SG, EP, RK, and RG contributed to the reviewing and editing. ET reviewed the manuscript.

Conflicts of Interest

RK is an employee of Novo Nordisk and serves as a director in Research and Development. RG is employed by Novo Nordisk and Disease Intelligence Pte Ltd and holds stock in both companies. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Missing data imputation.
[DOCX File, 23 KB - diabetes_v11i1e82635_app1.docx]

References

1. Diabetes. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/conditions-and-diseases/diabetes-and-other-endocrinal--nutritional-and-metabolic-conditions/diabetes [accessed 2026-01-19]
2. Pearson ER, Donnelly LA, Kimber C, et al. Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study. Diabetes 2007 Aug;56(8):2178-2182. [doi: 10.2337/db07-0440] [Medline: 17519421]
3. Uemura F, Okada Y, Torimoto K, Tanaka Y. Enlarged glycemic variability in sulfonylurea-treated well-controlled type 2 diabetics identified using continuous glucose monitoring. Sci Rep 2021 Mar 1;11(1):4875. [doi: 10.1038/s41598-021-83999-z] [Medline: 33649395]
4. Chung WK, Erion K, Florez JC, et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2020 Jul;43(7):1617-1635. [doi: 10.2337/dci20-0022] [Medline: 32561617]
5. Maddaloni E, Bolli GB, Frier BM, et al. C-peptide determination in the diagnosis of type of diabetes and its management: a clinical perspective. Diabetes Obes Metab 2022 Oct;24(10):1912-1926. [doi: 10.1111/dom.14785] [Medline: 35676794]
6. Garg S, Kitchen R, Gupta R, Pearson E. Applications of AI in predicting drug responses for type 2 diabetes. JMIR Diabetes 2025 Mar 27;10:e66831. [doi: 10.2196/66831] [Medline: 40146874]
7. Zou X, Liu Y, Ji L. Review: Machine learning in precision pharmacotherapy of type 2 diabetes—a promising future or a glimpse of hope? Digit Health 2023;9:20552076231203879. [doi: 10.1177/20552076231203879] [Medline: 37786401]
8. Hao Y, Cheng C, Li J, et al. Multimodal integration in health care: development with applications in disease management. J Med Internet Res 2025 Aug 21;27:e76557. [doi: 10.2196/76557] [Medline: 40840463]
9. Ren Q, Xiao D, Han X, et al. Genetic and clinical predictive factors of sulfonylurea failure in patients with type 2 diabetes. Diabetes Technol Ther 2016 Sep;18(9):586-593. [doi: 10.1089/dia.2015.0427] [Medline: 27403931]
10. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. Preprint posted online on May 22, 2017. [doi: 10.48550/arXiv.1705.07874]
11. Hébert HL, Shepherd B, Milburn K, et al. Cohort profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). Int J Epidemiol 2018 Apr 1;47(2):380-381j. [doi: 10.1093/ije/dyx140] [Medline: 29025058]
12. Martono DP, Lub R, Lambers Heerspink HJ, Hak E, Wilffert B, Denig P. Predictors of response in initial users of metformin and sulphonylurea derivatives: a systematic review. Diabet Med 2015 Jul;32(7):853-864. [doi: 10.1111/dme.12688] [Medline: 25582542]
13. Feng C, Wang H, Lu N, Tu XM. Log transformation: application and interpretation in biomedical research. Stat Med 2013 Jan 30;32(2):230-239. [doi: 10.1002/sim.5486] [Medline: 22806695]
14. Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics 2016 Sep 9;17(1):359. [doi: 10.1186/s12859-016-1236-x] [Medline: 27612635]
15. Van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011 Dec;45(3):67. [doi: 10.18637/jss.v045.i03]
16. Rubin DB. Multiple Imputation for Nonresponse in Surveys: John Wiley & Sons; 1987.
17. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007 Sep 11;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
18. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition: Springer Series in Statistics; 2009. [doi: 10.1007/978-0-387-84858-7]
19. García CB, Salmerón R, García C, García J. Residualization: justification, properties and application. J Appl Stat 2020;47(11):1990-2010. [doi: 10.1080/02664763.2019.1701638] [Medline: 35707574]
20. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32. [doi: 10.1023/A:1010933404324]

XSL•FO
RenderX

21. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. arXiv. Preprint posted online on Mar 9, 2016. [doi: 10.48550/arXiv.1603.02754]
23. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. Nature New Biol 2002;323:533-536. [doi: 10.1038/323533a0]
24. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. Ann Appl Stat 2010 Mar;4(1):266-298. [doi: 10.1214/09-AOAS285]
25. Kapelner A, Bleich J. bartMachine: machine learning with Bayesian additive regression trees. J Stat Soft 2016 Apr;70(4):40. [doi: 10.18637/jss.v070.i04]
26. Cardoso P, Young KG, Hopkins R, et al. Evaluating prediction of short-term tolerability of five type 2 diabetes drug classes using routine clinical features: UK population-based study. Diabetes Obes Metab 2025 Aug;27(8):4320-4329. [doi: 10.1111/dom.16470] [Medline: 40375735]
27. Cardoso P, Young KG, Nair ATN, et al. Phenotype-based targeted treatment of SGLT2 inhibitors and GLP-1 receptor agonists in type 2 diabetes. Diabetologia 2024 May;67(5):822-836. [doi: 10.1007/s00125-024-06099-3] [Medline: 38388753]
28. Drug response prediction. GitHub. 2025. URL: https://github.com/Pearson-Research-Lab/Drug-Response-Prediction [accessed 2026-01-13]
29. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79-82. [doi: 10.3354/cr030079]
30. Freedman D. Statistical Models: Theory and Practice: Cambridge University Press; 2005. [doi: 10.1017/CBO9781139165495]
31. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
32. Powers DMW. Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Mach Learn Technol 2008 Jan [FREE Full text]
33. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008 Nov;28(5):26. [doi: 10.18637/jss.v028.i05]
34. Jones AG, Lonergan M, Henley WE, Pearson ER, Hattersley AT, Shields BM. Should studies of diabetes treatment stratification correct for baseline HbA1c? PLoS ONE 2016;11(4):e0152428. [doi: 10.1371/journal.pone.0152428] [Medline: 27050911]
35. Dennis JM, Henley WE, Weedon MN, et al. Sex and BMI alter the benefits and risks of sulfonylureas and thiazolidinediones in type 2 diabetes: a framework for evaluating stratification using routine clinical and individual trial data. Diabetes Care 2018 Sep;41(9):1844-1853. [doi: 10.2337/dc18-0344] [Medline: 30072404]
36. Dennis JM, Young KG, Cardoso P, et al. A five-drug class model using routinely available clinical features to optimise prescribing in type 2 diabetes: a prediction model development and validation study. Lancet 2025 Mar 1;405(10480):701-714. [doi: 10.1016/S0140-6736(24)02617-5] [Medline: 40020703]

Abbreviations

AUC: area under the receiver operating characteristic curve
BART: Bayesian additive regression trees
C-peptide: connecting peptide
GoDARTS: Genetics of Diabetes Audit and Research in Tayside Scotland
HbA_{1c}: hemoglobin A_{1c}
HDL: high-density lipoprotein
ML: machine learning
NHS: National Health Service
NN: neural network
RMSE: root mean square error
SHAP: Shapley Additive Explanations
VIF: variance inflation factor
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by R Zhao; submitted 19.Aug.2025; peer-reviewed by J Low, M Yagcioglu, P Taiwo; revised version received 08.Dec.2025; accepted 11.Dec.2025; published 06.Feb.2026.

Please cite as:
Garg S, Kitchen R, Gupta R, Trucco E, Pearson E
Predictors of Glycemic Response to Sulfonylurea Therapy in Type 2 Diabetes Over 12 Months: Comparative Analysis of Linear Regression and Machine Learning Models
JMIR Diabetes 2026;11:e82635
URL: https://diabetes.jmir.org/2026/1/e82635
doi:10.2196/82635

© Shilpa Garg, Robert Kitchen, Ramneek Gupta, Emanuele Trucco, Ewan Pearson. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 6.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Privacy-Preserving Collaborative Diabetes Prediction in Heterogeneous Health Care Systems: Algorithm Development and Validation of a Secure Federated Ensemble Framework

Md Rakibul Hasan, MSc; Juan Li, PhD

Department of Computer Science, Faculty, North Dakota State University, 258 Quentin Burdick Bldg, Computer Science Department, NDSU, 1320 Albrecht Blvd, Fargo, ND, United States

Corresponding Author:
Juan Li, PhD
Department of Computer Science, Faculty, North Dakota State University, 258 Quentin Burdick Bldg, Computer Science Department, NDSU, 1320 Albrecht Blvd, Fargo, ND, United States

Abstract

Background: Diabetes prediction requires accurate, privacy-preserving, and scalable solutions. Traditional machine learning models rely on centralized data, posing risks to data privacy and regulatory compliance. Moreover, health care settings are highly heterogeneous, with diverse participants, hospitals, clinics, and wearables, producing nonindependent and identically distributed data and operating under varied computational constraints. Learning in isolation at individual institutions limits model generalizability and effectiveness. Collaborative federated learning (FL) enables institutions to jointly train models without sharing raw data, but current approaches often struggle with heterogeneity, security threats, and system coordination.

Objective: This study aims to develop a secure, scalable, and privacy-preserving framework for diabetes prediction by integrating FL with ensemble modeling, blockchain-based access control, and knowledge distillation. The framework is designed to handle data heterogeneity, nonindependent and identically distributed distributions, and varying computational capacities across diverse health care participants while simultaneously enhancing data privacy, security, and trust.

Methods: We propose a federated ensemble learning framework, FedEnTrust, that enables decentralized health care participants to collaboratively train models without sharing raw data. Each participant shares soft label outputs, which are distilled and aggregated through adaptive weighted voting to form a global consensus. The framework supports heterogeneous participants by assigning model architectures based on local computational capacity. To ensure secure and transparent coordination, a blockchain-enabled smart contract governs participant registration, role assignment, and model submission with strict role-based access control. We evaluated the system on the PIMA Indians Diabetes Dataset, measuring prediction accuracy, communication efficiency, and blockchain performance.

Results: The FedEnTrust framework achieved 84.2% accuracy, with precision, recall, and F_1-score of 84.6%, 88.6%, and 86.4%, respectively, outperforming existing decentralized models and nearing centralized deep learning benchmarks. The blockchain-based smart contract ensured 100% success for authorized transactions and rejected all unauthorized attempts, including malicious submissions. The average blockchain latency was 210 milliseconds, with a gas cost of ~107,940 units, enabling secure, real-time interaction. Throughout, patient privacy was preserved by exchanging only model metadata, not raw data.

Conclusions: FedEnTrust offers a deployable, privacy-preserving solution for decentralized health care prediction by integrating FL, ensemble modeling, blockchain-based access control, and knowledge distillation. It balances accuracy, scalability, and ethical data use while enhancing security and trust. This work demonstrates that secure federated ensemble systems can serve as practical alternatives to centralized artificial intelligence models in real-world health care applications.

(JMIR Diabetes 2026;11:e79166) doi:10.2196/79166

KEYWORDS

blockchain; decentralized health care; diabetes prediction; ensemble learning; federated learning; knowledge distillation; privacy-preserving AI; artificial intelligence; AI

Introduction

Diabetes continues to pose a growing global health burden, requiring timely prediction and proactive management to reduce complications and improve quality of life [1]. While machine learning has emerged as a powerful tool for diabetes prediction, conventional approaches often rely on centralized data repositories [2-4]. This reliance introduces serious challenges related to patient privacy, regulatory compliance (eg, Health Insurance Portability and Accountability Act (HIPAA), General

XSL•FO
RenderX

Data Protection Regulation (GDPR), and susceptibility to cyberattacks [5]. Moreover, centralized data aggregation is increasingly impractical due to fragmented data ownership across institutions and regions [6].

Real-world health care systems are inherently heterogeneous, encompassing a wide range of contributors—from large hospitals and urban clinics to wearable health devices in remote settings [7]. These entities vary significantly in data volume, quality, and computational capacity. The data are often nonindependent and identically distributed (non-IID), reflecting demographic, clinical, and behavioral diversity [8]. As a result, models trained within a single institution or on homogeneous datasets often struggle to generalize across settings, limiting their effectiveness and scalability.

To address these limitations, collaborative federated learning (FL) has emerged as a compelling solution [9]. However, applying FL to real-world diabetes prediction presents several unresolved challenges. In particular, current FL frameworks often struggle with:

- security vulnerabilities, such as model poisoning and adversarial manipulation [10]
- lack of coordination and trust, especially in decentralized, multiparty settings [11]
- performance degradation due to client heterogeneity and non-IID data distributions [12]

While several FL frameworks [13-16] have been explored for decentralized health care analytics, most assume homogeneous model architectures, single global models, or idealized trust environments and do not explicitly address lightweight or resource-constrained participants at the edge [17,18]. Existing systems, such as Biscotti [19] and Chang et al [20], rely on gradient sharing and therefore require structurally aligned models and consistent computational resources, while recent blockchain-enabled FL frameworks incorporate differential privacy but still assume homogeneous models or centralized coordination [21,22]. Furthermore, blockchain [23], a promising technology for ensuring integrity, transparency, and access control in decentralized systems, has seen limited integration with FL, especially in diabetes prediction contexts. Other blockchain-enabled approaches, such as Shalan et al [24], provide secure access control but do not incorporate mechanisms for interoperable knowledge sharing across heterogeneous local models.

In contrast, FedEnTrust introduces an integrated design that simultaneously addresses model heterogeneity, non-IID data, trust and identity verification, and secure update submission. By combining soft-label knowledge distillation with blockchain-verified RBAC, FedEnTrust enables robust collaboration across diverse health care systems while preventing unauthorized or malicious updates. FedEnTrust introduces a novel integration of:

- ensemble learning, allowing clients to train diverse local models best suited to their data and computational constraints
- soft-label knowledge distillation, enabling effective model aggregation across non-IID participants
- blockchain-based smart contracts, which provide tamper-proof coordination, role-based access control, and participant accountability

FedEnTrust represents a step forward in secure and collaborative artificial intelligence (AI) for health care, with the following key contributions:

1. Heterogeneity-aware ensemble design: Each participant trains a model tailored to its resource level, supporting real-world deployment across varied health care nodes.
2. Knowledge distillation-based aggregation: We introduce a soft-label ensemble mechanism that improves convergence and generalization across non-IID data.
3. Blockchain-enabled trust layer: Our smart contract system enforces participant registration, access control, and secure model submissions without a centralized authority.
4. Comprehensive evaluation: Using the PIMA Indians Diabetes Dataset, we demonstrate that FedEnTrust improves prediction accuracy; maintains privacy; and ensures secure, low-latency collaboration.

By addressing the intersection of privacy, trust, heterogeneity, and security, FedEnTrust provides a practical and deployable framework for AI-powered diabetes prediction in real-world, decentralized health care systems.

Methods

Overview of FedEnTrust

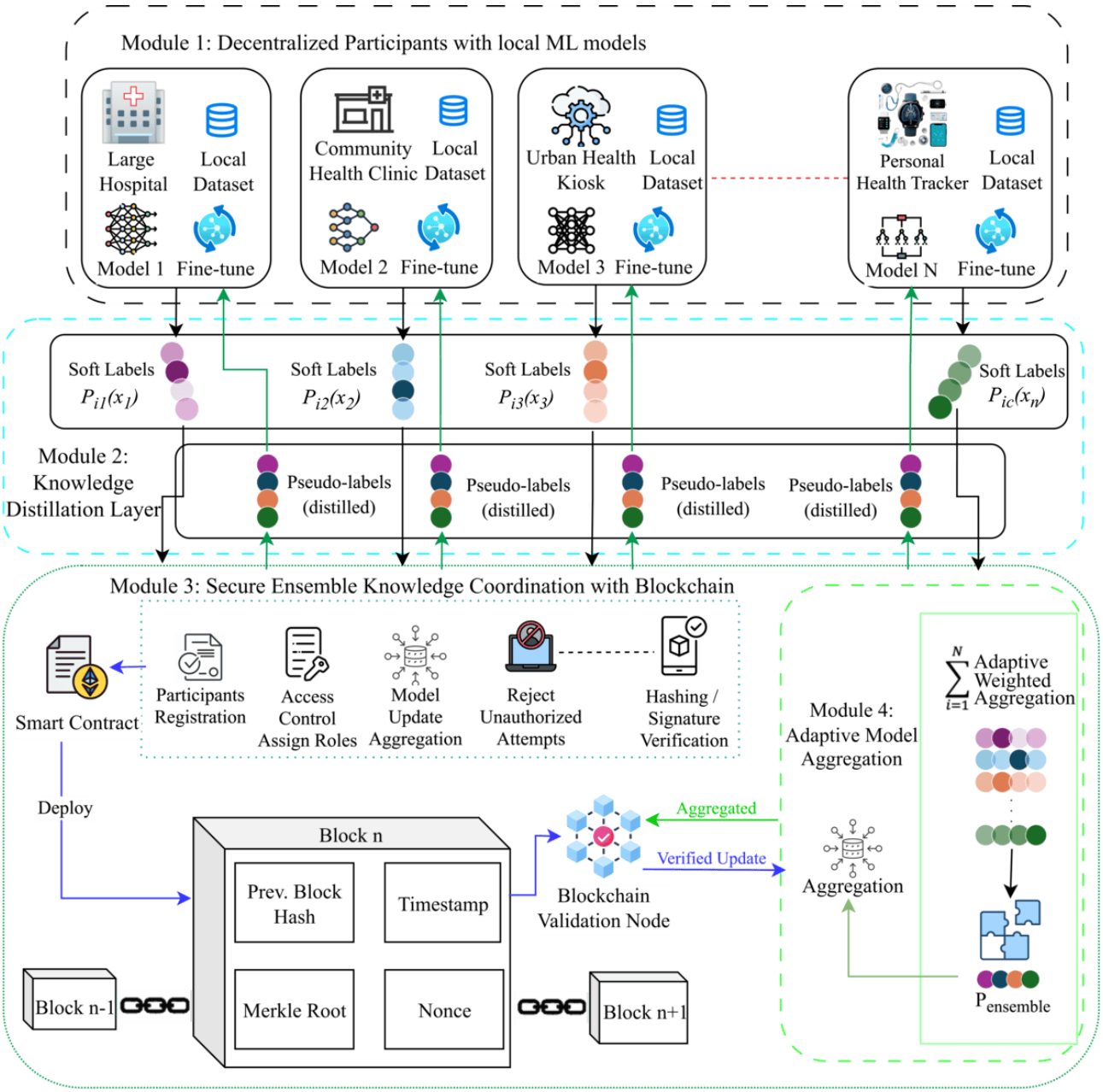
FedEnTrust is a secure, privacy-preserving federated ensemble learning framework designed to address the challenges of decentralized diabetes prediction across heterogeneous health care environments. It enables collaborative learning without centralizing sensitive patient data, accommodates diverse computational resources, and defends against malicious behaviors through a blockchain-coordinated trust infrastructure. The core modules of FedEnTrust include (1) heterogeneity-aware local model training, (2) knowledge distillation via soft label sharing, (3) blockchain-based secure coordination, and (4) adaptive ensemble aggregation.

These modules work together to realize 3 key objectives: maintaining patient privacy, enabling equitable participation across institutions with varying capacities, and ensuring secure collaboration in a decentralized system.

Figure 1 illustrates the end-to-end data flow across the 4 modules. Local raw data remain strictly on the device. Each participant trains a heterogeneous local model and generates soft-label probability vectors. These soft labels, along with accuracy metadata, are sent off-chain to the aggregator but must first pass through blockchain-based role-based access control (RBAC) validation, where the smart contract verifies participant identity, role permissions, and submission metadata. Validated soft labels are incorporated into an adaptive weighted aggregation mechanism, producing global pseudo-labels that are redistributed to all participants. The blockchain records transaction hashes and role enforcement events, ensuring traceability without revealing sensitive data.

XSL•FO
RenderX

Figure 1. Overview of the FedEnTrust architecture. Soft labels generated by local models are authenticated through blockchain-based role-based access control and combined using adaptive weighted aggregation to produce pseudo labels for continued local training. ML: machine learning.



Architectural Novelty and Comparison With Existing FL Frameworks

Real-world health care environments exhibit substantial diversity in computational capacity, data distributions, trust requirements, and security risks. To contextualize the design of FedEnTrust within this landscape, we compare its architectural capabilities against representative FL and blockchain-enabled frameworks in Table 1.

XSL•FO
RenderX

Table . Architectural comparison of FedEnTrust with representative federated learning frameworks.

Challenge in real-world health care FL	FedEnTrust (Our work)	Hasan et al [15]	Biscotti [19]	Chang et al [20]	Microcontroller FL[a] [17]
Heterogeneous compute environments (hospitals, clinics, kiosks, wearables)	Heterogeneity-aware model assignment; each node trains model matching its device capacity; ensemble aggregation aligns knowledge across disparate models	Supports ML[b] models but generally assumes similar capacity clients	Assumes all clients run comparable gradient-sharing deep models	Single model structure required; difficult for low-resource clients	Designed for ultra-low-power devices; not suitable for multitier health care
Non-IID[c] and imbalanced data across institutions	Soft-label knowledge distillation + weighted aggregation improve cross-site generalization	Local models trained independently; static averaging struggles with non-IID distributions	Gradient aggregation without distillation; non-IID data reduces convergence	DP[d]-sanitized gradients reduce signal strength on non-IID data	Very limited support for complex non-IID medical data
Cross-institution trust and secure participation	Smart contract–driven RBAC[e]; on-chain validation of model submissions; rejects malicious or unauthorized updates	Minimal security; no on-chain validation	Uses blockchain only as consensus layer, not for role-level access control	Smart contract manages DP gradients, not participation permissions	No trust or participation assurance mechanism
Protection against malicious updates (poisoning, fake uploads)	On-chain validator roles + metadata checks prevent poisoned soft labels before aggregation	No defense against malicious gradient or model uploads	Consensus prevents tampering but not model poisoning	DP reduces leakage but not poisoning	No adversarial defense features
Interoperability across model types	Soft labels unify outputs of RF[f], XGB[g], DT[h], SVM[i], KNN[j] into comparable probability space	Homogeneous ML models; limited interoperability	Requires same model structure for gradient fusion	Single-model FL; weights must match	No model interoperability
Scalability across distributed health care networks	Lightweight soft-label sharing reduces communication overhead and suits mixed-resource environments	Local model averaging; moderate scalability	Heavy blockchain consensus overhead limits scalability	DP gradient exchange increases bandwidth needs	Limited to microcontroller networks
Auditability and traceability for compliance (HIPAA[k] or GDPR[l])	Full on-chain audit log of registrations, updates, and permissions	Centralized coordination; limited auditability	All gradient updates stored on-chain—high cost	Stores only gradient summaries; limited audit transparency	Not designed for regulated health care settings

[a]FL: federated learning.
[b]ML: machine learning.
[c]IID: independent and identically distributed.
[d]DP: differential privacy.
[e]RBAC: role-based access control.
[f]RF: random forest.
[g]XGB: extreme gradient boosting.
[h]DT: decision tree.
[i]SVM: support vector machine.
[j]KNN: k-nearest neighbors.
[k]HIPAA: Health Insurance Portability and Accountability Act.
[l]GDPR: General Data Protection Regulation.

Unlike approaches such as Hasan et al [15], Biscotti [19], and Chang et al [20], which rely on homogeneous model structures or gradient-based updates, FedEnTrust supports heterogeneity-aware model assignment. Each participant trains a locally suitable model (eg, random forest, extreme gradient boosting, decision tree, support vector machine [SVM], k-nearest neighbors [KNN]) based on its available resources, enabling participation from hospitals, clinics, kiosks, and wearable devices.

FedEnTrust also differs from blockchain-enabled systems such as Shalan et al [24] and TinyFL [25]. While these frameworks

XSL•FO
RenderX

integrate blockchain for logging or access control, they do not incorporate soft-label knowledge distillation or adaptive ensemble aggregation to unify heterogeneous model outputs. FedEnTrust introduces a unique coupling of soft-label–based distillation with blockchain-enforced RBAC, enabling secure verification of participant identity and role prior to model update submission, on-chain logging of update hashes to ensure auditability, prevention of malicious or unauthorized contributions before they influence aggregation, and interoperability of predictions across diverse model architectures.

This integration ensures that only authenticated, validated soft labels contribute to the global model. This design is particularly effective for non-IID and imbalanced health care data settings, where traditional gradient-averaging approaches struggle.

Module 1: Decentralized Local Training With Heterogeneous Models

FedEnTrust begins with a network of decentralized health care participants, including large hospitals, regional clinics, kiosks, and personal health trackers, each training its own machine learning model locally. These models are tailored to each participant's computational capabilities and data volume. For example, high-resource hospitals may use deep neural networks, while low-resource settings use shallow learning such as KNN or support vector classifier (SVC) to support real-time inference with minimal memory demands.

This heterogeneity-aware model assignment ensures that all participants, regardless of scale or technical capacity, can contribute meaningfully. Local training is performed privately using internal datasets, aligning with privacy regulations such as HIPAA and GDPR.

Module 2: Knowledge Distillation via Soft Labels

To facilitate collaborative learning without exposing raw data, participants generate soft labels, probability distributions over prediction classes (eg, diabetic, nondiabetic). These soft labels encode richer information than binary outputs and are shared with a central aggregator, enabling cross-site knowledge transfer.

Soft Label Generation

Each participant generates soft labels, probability distributions reflecting its model's confidence across classes, and transmits these predictions to the aggregator. Unlike gradient-based approaches, soft labels create an interoperable representation across heterogeneous model types. Before being used for ensemble aggregation, every soft label submission is paired with metadata including local validation accuracy, model identifier, and round number. For an input instance *x*, the participant's model outputs a probability vector:

$$(1)\ Pi(x)=[p1,p2,\ldots,pc]\in Rc,\ \text{where}\ \sum c=1Cpc=1$$

These soft labels encapsulate the model's confidence across the Cclasses and support knowledge transfer without sharing raw patient data or internal model parameters.

To address differences in how heterogeneous models calibrate probability outputs, FedEnTrust applies temperature scaling, which smooths the probability distribution by dividing logits zi(x)by a temperature parameter T:

$$(2)\ Pi(t)(x)=\text{softmax}\ (zi(x)T),\ T=2$$

A temperature of T=2 was selected because values greater than 1 produce smoother, less overconfident probability distributions, which improves the stability of aggregation across models with different calibration characteristics. A small temperature (eg, T=1) can lead to overly sharp probabilities that amplify noise, while excessively large values dilute useful predictive signals. Empirical testing showed that T=2 offers an optimal balance.

Dynamic Weight Updates Across Federated Rounds

Once soft labels are generated by each participant model, the system proceeds to combine these distributed outputs into a unified global prediction. This ensemble consensus represents a key step in transferring collective intelligence across all nodes while respecting the constraints of data privacy and computational diversity.

The ensemble aggregation process employs adaptive weighted soft voting, where more reliable and accurate models are given stronger influence. For example, a well-resourced clinic with consistently high validation performance will contribute more to the global prediction than a basic kiosk with limited data. However, no participant is excluded; each contributes according to its validated strength, ensuring fairness and inclusivity in the learning process. FedEnTrust adaptively updates the influence of each participant during communication round t. Each participant evaluates its model using a shared public validation subset to compute Accit, which is the validation accuracy of participant i at round t. The ensemble assigns each participant a normalized contribution weight:

$$(3)\ Wi(t)=Acci(t)\sum j=1NAccj(t)$$

To prevent dominant institutions (eg, large hospitals with more data) from exerting disproportionate influence, FedEnTrust applies weight clipping, capping Witat an upper bound. This ensures contribution fairness and reduces the risk of bias toward specific demographic subpopulations.

Justification for Heterogeneous Model Assignment

The model architectures listed in Table 2 were selected to reflect realistic resource constraints and deployment contexts:

- Random forest (hospitals): Hospitals possess sufficient computational capacity and large datasets; random forest models capture nonlinear relationships and perform well on tabular clinical data.
- XGB (regional clinics): XGB provides strong performance under moderate computational resources, making it suitable for mid-sized clinics.
- Decision trees and KNN (community clinics or kiosks): These models require minimal training cost and support real-time inference in low-power environments.
- Linear SVM (wearables or personal trackers): Linear SVM has a lower memory footprint than logistic regression and offers more stable performance on small, noisy physiological samples typically produced by wearables.

XSL•FO
RenderX

Table . Simulated decentralized participants and their models.

ID	Participant	Model architecture	Key parameters	Resource level	Weight	Remarks
1	Large hospital	Random forest	n_estimators=130max_depth=15max_features=0.75 data_use=50%	Very high	0.50	Trains complex models on large datasets; serves as a high-capacity node
2	Urban health kiosk	K-nearest neighbors	n_neighbors=5 algorithm='auto'data_use=5%	Low	0.05	Designed for low-resource environments using simple, efficient models
3	Regional clinic	XGBoost	learning_rate=0.01max_depth=10n_estimators=180data_use=30%	High	0.30	Supports moderately complex modeling on medium-sized datasets
4	Community health clinic	Decision tree	max_depth=Nonecriterion='gini'data_use=10%	Medium	0.10	Runs interpretable tree-based models with moderate resource needs
5	Personal health tracker	Support vector machine	kernel='linear'C=1.0data_use=5%	Very low	0.05	Uses lightweight models suitable for wearables and embedded devices

This heterogeneity-aware mapping allows each participant to train a model aligned with its resource profile while still contributing to a unified ensemble.

Enhanced Knowledge Distillation and Pseudo-Label Generation

In each communication round t, participant models generate calibrated soft probability vectors Pit(x), which are aggregated using dynamically updated participant weights to produce a global soft prediction.

Our proposed model aggregates the calibrated soft labels using the dynamic weights to produce a global soft prediction:

$$(4)Pt(x)=\sum i=1NWit*Pit(x)$$

Because aggregation operates entirely on probability distributions rather than gradients or model parameters, FedEnTrust naturally supports heterogeneous machine learning architectures across hospitals, clinics, kiosks, and personal wearable devices while preserving data locality and privacy.

To improve the reliability of knowledge transfer, each participant's soft predictions undergo normalization followed by temperature scaling (with T=2) to smooth overconfident outputs. The ensemble output is then evaluated using a confidence-based filtering mechanism, where pseudo-labels are generated only if the maximum ensemble probability satisfies:

$$(5)\max(Pt(x))\geq\tau$$

With τ=0.7 Predictions failing this criterion are discarded to prevent the propagation of uncertainty or noise. Accepted pseudo-labels are normalized and redistributed to participants, where they are appended to local datasets and used for continued training in the subsequent round. This feedback loop enables low-resource participants to benefit from globally distilled knowledge while retaining local autonomy.

All soft-label submissions are validated through the blockchain-based RBAC mechanism described in Module 3. Only soft labels originating from authenticated and authorized roles (eg, model-provider) are accepted. Validated submissions are incorporated into an adaptive weighted soft-voting process, where participant weights are updated based on observed local performance across rounds. The resulting global outputs are then redistributed as pseudo-labels for the next training iteration, ensuring robustness against non-IID data distributions, preventing malicious or fabricated updates, and enhancing cross-site generalization across heterogeneous health care environments.

Module 3: Blockchain-Based Secure Coordination

Overview

Module 3 employs an Ethereum-based smart contract to authenticate participants, enforce role permissions, and log immutable update metadata. When a node attempts to upload soft labels, the smart contract verifies the participant's role, identity, timestamp, and declared accuracy. The contract then generates and stores a hashed representation of the update, which validator nodes review. Only soft labels that receive approval from multivalidators are admitted to the aggregation pool. This ensures tamper resistance, prevents poisoning attacks, and provides end-to-end traceability for health care compliance requirements. When a participant attempts to contribute soft labels, the smart contract performs the following checks:

1. Identity verification: Confirms that the contributor is a registered network participant.
2. Role validation: Ensures the contributor holds a permitted role to submit model outputs.
3. Metadata verification: Confirms the integrity of reported metrics (eg, accuracy, round number).

XSL•FO
RenderX

4. Hash logging: Stores a transaction hash to provide auditability without exposing any data.

Only after passing these checks is the soft label included in the aggregation pool. This design prevents poisoned or fabricated updates from influencing the global model and eliminates single points of failure in participation management. By integrating RBAC directly with knowledge distillation, FedEnTrust establishes a secure and transparent trust layer that coordinates collaborative learning across diverse health care nodes.

Blockchain Platform Selection and Justification

FedEnTrust is implemented on an Ethereum-compatible private blockchain network. Ethereum was selected due to its deterministic smart contract execution, robust security guarantees, and mature tooling ecosystem. The platform supports Solidity-based smart contracts, Remix IDE integration, and widely adopted standards for access control and event logging. These characteristics make Ethereum well suited for privacy-preserving health care collaboration, where verifiable execution and auditability are required.

To justify this choice, we compared Ethereum with 2 commonly used permissioned blockchain platforms: Hyperledger Fabric and Corda. Table 3 presents a feature-level comparison of Ethereum, Hyperledger Fabric, and Corda across network type, decentralization, smart contract support, privacy mechanisms, ecosystem maturity, and application alignment. Given the need for flexible smart contract logic, verifiable coordination, and broad compatibility with Internet of Things (IoT) and health care prototypes, Ethereum provides the most practical platform for FedEnTrust.

Table . Comparison of blockchain platforms.

Feature	Ethereum	Hyperledger fabric	Corda
Network type	Public or private	Permissioned	Permissioned
Decentralization	Highly decentralized	Semi-decentralized	Semi-decentralized
Smart contracts	Solidity, robust tooling	Chaincode (Go/Java/Node.js)	Contract flows for financial logic
Privacy	Extensible via Layer-2/private networks	Strong privacy (channels, private collections)	Strong bilateral privacy
Ecosystem	Very large developer ecosystem	Enterprise-focused	Financial institutions
Use alignment	Decentralized coordination across heterogeneous nodes	Consortium-style enterprise networks	Regulated financial workflows

Adversarial Threat Model and Security Resilience

FL deployments in real-world health care environments may be exposed to adversarial participants attempting to manipulate the global model, disrupt training, or infer sensitive information. To address these risks, we construct a structured threat model covering three primary attack categories: (1) model poisoning; (2) collusion among compromised participants; and (3) malicious soft-label injection, where adversaries submit manipulated pseudo-probabilities to bias the aggregation process.

FedEnTrust incorporates multiple, tightly coupled defense mechanisms across its blockchain coordination and ensemble aggregation layers to provide resilience against these threats.

1. Model poisoning and malicious soft-label injection: A compromised participant may attempt to submit adversarial or fabricated soft labels to influence global predictions. FedEnTrust mitigates this risk through smart contract–enforced RBAC, which restricts update submission exclusively to authenticated participants holding an authorized model-provider role. Each submission is accompanied by metadata including round number, reported validation accuracy, and timestamp, which are verified for internal consistency before acceptance. To ensure integrity and prevent replay or tampering, all submissions are cryptographically hashed and logged on-chain. Furthermore, FedEnTrust employs validator redundancy, requiring approval from multiple trusted validator nodes (eg, lead hospitals within the consortium) before a submission is incorporated into aggregation, preventing single-node compromise.
2. Collusion and validator compromise: To reduce the impact of colluding or compromised participants, FedEnTrust adopts a consortium-style multivalidator approval mechanism. No single validator can independently approve a model update; instead, a quorum of validators must jointly authorize submissions. The validator set itself is managed through governed smart contract functions, allowing secure updates to validator membership over time and eliminating static trust assumptions.
3. Blockchain-specific threats: Public blockchain deployments may be vulnerable to front-running, transaction reordering, or gas manipulation attacks. FedEnTrust avoids these risks by operating on a private Ethereum-compatible consortium network without a public mempool, eliminating front-running opportunities. Smart contracts use fixed gas budgets and sequential transaction counters to ensure deterministic execution and prevent reordering attacks.
4. Privacy leakage through on-chain metadata: Although raw data and model parameters are never shared, metadata leakage can still pose privacy risks. FedEnTrust minimizes exposure by storing only hashed identifiers and role-verification logs on-chain. No patient-level attributes, raw predictions, or model parameters are recorded. All soft labels remain strictly off-chain and are exchanged only between authorized participants and the aggregator over secure channels.
5. Aggregation-level safeguards: Beyond blockchain enforcement, the adaptive ensemble layer further mitigates

XSL•FO
RenderX

adversarial influence by applying temperature scaling, confidence thresholds, and weight clipping. These mechanisms limit the amplification of extreme or adversarial soft-label probabilities and restrict the maximum influence any single participant can exert, even if it reports high accuracy.

Collectively, these mechanisms establish a multilayered security architecture that protects FedEnTrust against common poisoning, collusion, and manipulation attempts at the coordination and authorization layers while preserving decentralized operation and data privacy. The empirical results demonstrate that unauthorized and malicious submissions are consistently detected and rejected through blockchain-enforced RBAC and validator checks. While this study focuses on secure enforcement and system robustness rather than controlled adversarial learning simulations, the framework is explicitly designed to support future evaluation against targeted and untargeted attacks, including label-flipping, probability-shifting, and adaptive adversarial strategies.

Module 4: Adaptive Model Aggregation and Feedback Loop

After soft labels are aggregated into a global ensemble prediction, FedEnTrust redistributes this consensus to participants as pseudo-labels for retraining. This adaptive aggregation ensures that high-performing models contribute more to the global prediction, while low-resource nodes still benefit from the collective knowledge.

This module enables faster convergence across non-IID data, fair and inclusive participation, and improved generalization without data sharing.

The result is a balanced feedback loop: local models become more aligned with the ensemble, improving personalization and global performance over time.

System Implementation and Evaluation Setup

We evaluated FedEnTrust using the publicly available PIMA Indians Diabetes Dataset [26], which includes 768 records of female patients with 8 clinical attributes and a binary diabetes outcome. Data were preprocessed using the following steps:

1. Outlier detection with IQR and local outlier factor
2. Feature engineering (eg, binning glucose, insulin levels)
3. Normalization using *z* scores
4. Class balancing using the synthetic minority oversampling technique [27]

As shown in Table 1, to simulate a real-world heterogeneous environment, the dataset was split across 5 simulated participants with varying data volumes and models. Each participant's computational weight was reflected in the aggregation process, mimicking operational conditions ranging from large hospitals to low-power personal devices.

Ethical Considerations

This study exclusively used publicly available, deidentified secondary datasets. No new data were collected, and no interaction with human participants occurred. According to institutional policy and US federal regulations (45 CFR 46), research involving publicly available, deidentified data does not constitute human participant research and is therefore exempt from institutional review board review. As a result, institutional review board approval was not sought, and informed consent was not required. All datasets used in this study were fully deidentified prior to public release. The data contained no direct or indirect identifiers, and no attempt was made to reidentify individuals. Data were accessed and analyzed in accordance with the terms and conditions specified by the data providers. No participants were recruited for this study, and no compensation was provided.

Results

Model Performance

We evaluated the FedEnTrust framework across 5 heterogeneous participants over 15 communication rounds, focusing on prediction accuracy, precision, recall, and F_1-score. The results highlight how collaborative learning and adaptive aggregation significantly enhance performance, especially for participants with limited data and computational resources.

Figure 2 shows the accuracy trajectories of each participant over the FL rounds. Participant 1 (random forest), equipped with the largest dataset and the highest computational power, consistently achieved the highest accuracy, acting as a de facto "teacher" during knowledge distillation. Its influence helped guide improvements in lower-resource nodes, such as participant 5 (SVC) and participant 2 (KNN), which showed steady gains over time.

XSL•FO
RenderX

Figure 2. Global model accuracy improves over ensemble federated round. DT: decision tree; KNN: k-nearest neighbors; RF: random forest; SVC: support vector classifier; XGB: extreme gradient boosting.

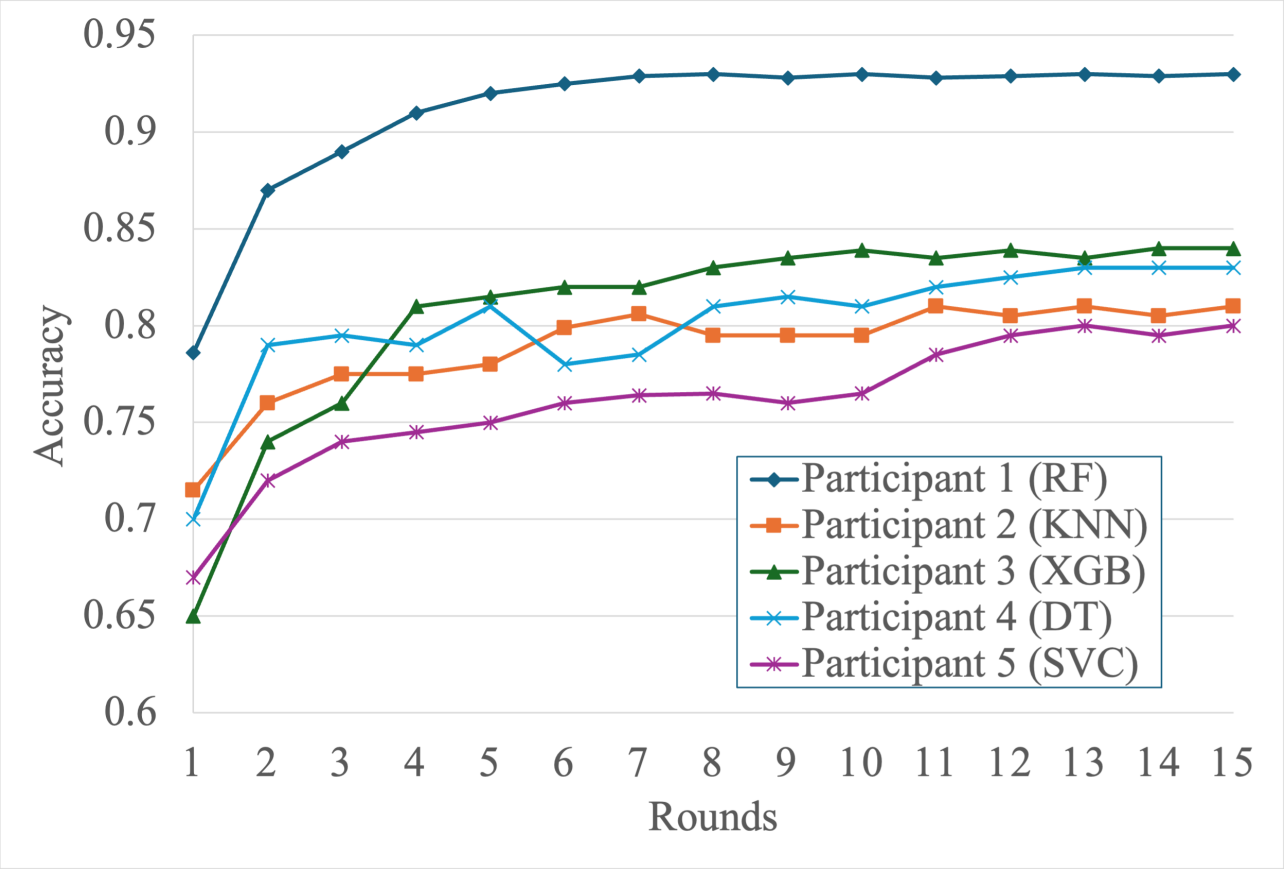


Figure 3 presents the corresponding model loss curves. All participants experienced substantial loss reduction early on, with convergence observed by round 15. Participant 1 maintained the lowest loss throughout, while participants 4 and 5 showed marked improvement from higher initial losses, demonstrating the benefit of federated collaboration.

XSL•FO
RenderX

Figure 3. Federated model losses over rounds. DT: decision tree; KNN: k-nearest neighbors; RF: random forest; SVC: support vector classifier; XGB: extreme gradient boosting.

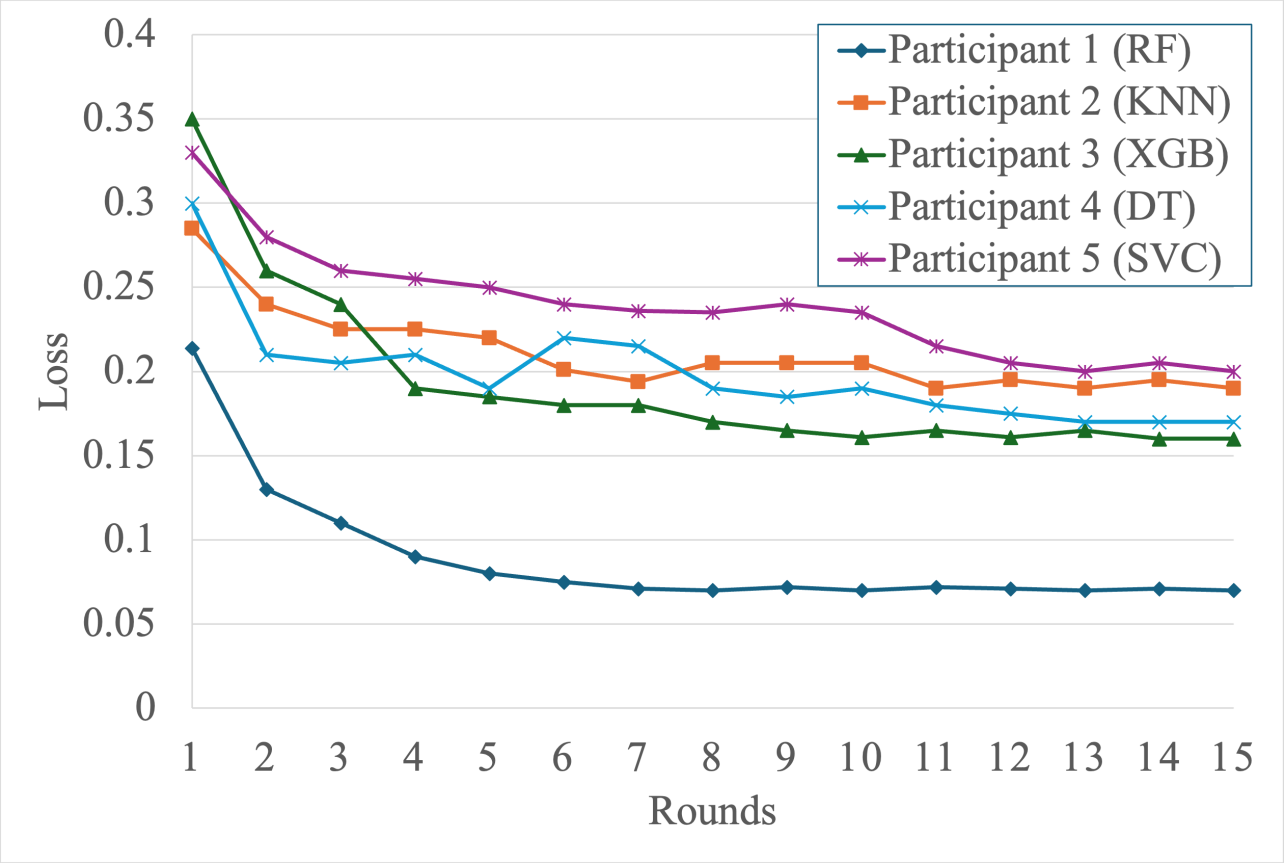


Table 5. Federated Models Performance after 15 rounds

Comparing the initial and federated performance results (Tables 4 and 5) reveals substantial gains for all participants after collaborative training. Accuracy improvements of up to 28% are observed in lower-resource participants, and F_1-scores increase consistently across all models, demonstrating the effectiveness of knowledge distillation and adaptive aggregation in heterogeneous environments. For example, participant 4 (decision tree) improves its F_1-score from 0.71 to 0.88, while participant 3 (XGBoost) improves from 0.64 to 0.85, highlighting the benefits of ensemble-driven knowledge transfer.

Table . Initial models' performance.

Participant	Accuracy	Precision	Recall	F_1-score
1	0.78	0.85	0.84	0.83
2	0.71	0.73	0.71	0.72
3	0.65	0.63	0.65	0.64
4	0.70	0.73	0.71	0.71
5	0.67	0.67	0.68	0.67

Table . Federated models' performance after 15 rounds.

Participant	Accuracy	Precision	Recall	F_1-score
1	0.93	0.92	0.94	0.93
2	0.81	0.80	0.86	0.83
3	0.84	0.85	0.86	0.85
4	0.83	0.87	0.90	0.88
5	0.80	0.79	0.87	0.83

To further characterize performance stability across communication rounds, Table 6 reports both the final accuracy at round 15 and the mean (SD) of accuracy over all 15 federated rounds. The relatively low SDs indicate stable convergence

XSL•FO
RenderX

behavior for all participants, even for lightweight models such as KNN and SVC. These results confirm that FedEnTrust effectively accommodates device and data heterogeneity while maintaining strong predictive performance, privacy preservation, and decentralized operation. Tailored model architectures, aligned with participant resource constraints, ensure balanced contribution and efficient deployment across the collaborative learning process.

Table . Federated model accuracy and variability across 15 rounds.

Participant	Model	Final accuracy	Accuracy, mean (SD)
1	RF[a]	0.93	0.91 (0.04)
2	KNN[b]	0.81	0.79 (0.03)
3	XGB[c]	0.84	0.81 (0.05)
4	DT[d]	0.83	0.80 (0.03)
5	SVC[e]	0.80	0.76 (0.03)

[a]RF: random forest.

[b]KNN: k-nearest neighbors.

[c]XGB: extreme gradient boosting.

[d]DT: decision tree.

[e]SVC: support vector classifier.

To assess whether the performance differences between FedEnTrust and baseline models were statistically meaningful on the PIMA Indians Diabetes Dataset, we conducted a nonparametric bootstrap significance analysis using the same held-out test set as the main evaluation. Because accuracy, precision, recall, and F_1-score are bounded metrics that may deviate from normality, bootstrap resampling provides a distribution-free and robust alternative to parametric methods such as the *t* test. We used a 2-tailed *t* test, as no directional assumption was imposed a priori and the objective was to assess whether there was any statistically significant difference between the compared methods.

We generated B=1000 bootstrap resamples by sampling test instances with replacement from the held-out evaluation set. For each bootstrap resample, we evaluated FedEnTrust and the decentralized baseline from Blockchain-FL with Differential Privacy [20], which represents the closest methodologically comparable prior work under similar privacy and decentralization constraints. This procedure produced 1000-sample empirical distributions for both models' accuracy. To quantify comparative performance, we computed the bootstrap metric difference for each resample:

$$\Delta(b)=M_{FedEnTrust}(b)-M_{Baseline}(b) \quad (6)$$

where Mb represents the accuracy, precision, recall, or F_1-score on bootstrap resample b. We then constructed 95% CIs for each metric difference using the percentile method.

The bootstrap CI analysis indicates that FedEnTrust achieves statistically significant performance improvements over the decentralized blockchain-based FL baseline [20]. Specifically, FedEnTrust attains a mean accuracy of 0.842 with a 95% bootstrap CI of 0.831-0.853, compared to 0.827 (0.814-0.839) for the decentralized baseline. The resulting accuracy difference of +0.015 yields a 95% CI of 0.004-0.027, which excludes zero, indicating statistical significance at α=.05. These results confirm that the performance gains observed for FedEnTrust are not due to random variation but rather stem from its integration of heterogeneous ensemble learning with blockchain-backed coordination under privacy constraints.

These findings validate that FedEnTrust's performance gains are not only empirical but statistically robust, reinforcing the effectiveness of combining heterogeneous ensemble learning with blockchain-backed coordination in constrained health care environments.

Blockchain Performance

We deployed the smart contract with 6 key functions and evaluated it under a realistic configuration consisting of 5 decentralized health care participants and 1 global aggregator. These components facilitated secure collaboration, access management, and federated training. The details are shown in Table 7.

Table . Blockchain system configuration.

Operation	Count	Description
Total registered participants	5	Registered using registerClient()
Federated coordination nodes	1	Global aggregator for accuracy aggregation and model ensemble
Smart contract functions deployed	6	Includes registration, role assignment, update logging, and access checks

XSL•FO
RenderX

To assess computational efficiency, we monitored key metrics such as gas consumption, data size, and latency for major smart contract operations. These measurements reflect the cost-effectiveness and responsiveness of blockchain-mediated tasks.

These operations incur gas overhead beyond Ethereum's 21,000 base gas due to additional computation, state updates, and event emissions. The *modelUpdate*() function, for example, consumes about 98,560 gas (~295 bytes of encoded parameters), balancing cost with functional depth and traceability (Table 8).

Table . Smart contract performance metrics.

Operation	Average gas cost	Data size (bytes)	Average latency (ms)
Client registration	118,073	370	220
Role assignment	109,820	345	210
Model update	98,560	295	195
Model aggregation	105,310	315	215

Despite slight delays compared to traditional systems, the observed latency (195 - 220 ms) remains acceptable for health care applications, considering the gains in trust, verifiability, and tamper resistance. To assess longer-term stability, we analyzed all 212 smart contract operations recorded during the training. All valid transactions executed successfully without anomalies, indicating stable performance across repeated interactions. The expanded evaluation in Table 9 includes average latency, latency range, and variability across extended cycles. These findings support the suitability of the blockchain layer for multiround federated training.

Table . Transaction integrity and enforcement metrics.

Category	Values	Description
Total transactions	212	All smart contract operations
Valid transactions	201	Successfully executed by authorized participants
Rejected transactions	11 (5.19%)	Unauthorized queries (6), malicious submissions (3), invalid role updates (2)
Success rate	100%	All valid transactions completed without error
Average latency	21.4 ms	Mean execution time for valid transaction
Latency range	14.8 - 36.2 ms	Minimum and maximum observed latency
SD	±4.7 ms	Variability in execution time
Latency over extended cycles (100 iterations)	Mean: 22.1 ms; variation:±5.3 ms	Long-term stability testing simulating multiround FL[a]
Finality time	~1 block(~1 s)	Deterministic finality in private PoA[b] Ethereum network
Estimated throughput	~47 tx/s	Consistent with private Ethereum networks

[a]FL: federated learning.

[b]PoA: proof-of-authority.

As illustrated in Figure 4, unauthorized model submissions are automatically rejected, triggering an on-chain error: "Client not registered." This ensures that only authenticated nodes contribute to the learning process, strengthening data integrity.

XSL•FO
RenderX

Figure 4. Access rejection for unauthorized participant.

Throughout 15 communication rounds, the smart contract reliably supported secure, real-time exchange of soft label predictions and model aggregation updates. For instance, participant 1 improved from 78% to 93% accuracy, while participant 4 rose from 70% to 83%, all while maintaining privacy and resisting tampering.

These results underscore the effectiveness of combining blockchain with federated ensemble learning to achieve scalable, secure, and privacy-preserving AI in health care environments.

Discussion

Principal Findings

This study presents FedEnTrust, a blockchain-enabled federated ensemble learning framework that offers a privacy-preserving and scalable solution for decentralized diabetes prediction. Our system effectively balances accuracy, privacy, and adaptability by integrating diverse machine learning models with knowledge distillation and adaptive weighted aggregation. With a predictive accuracy of 84.2%, FedEnTrust demonstrates competitive performance while maintaining strict privacy guarantees and supporting heterogeneous health care participants ranging from hospitals to wearable devices.

The framework's integration with blockchain smart contracts provides secure participant coordination, role-based access control, and transparent model validation without incurring substantial latency or resource overhead. Importantly, our results show that even low-resource participants benefit from collaboration through soft label exchange, enabling equitable participation in the learning process.

Comparison With Prior Work

Table 10 summarizes the performance of FedEnTrust against the existing centralized and decentralized methods applied to the PIMA Indians Diabetes Dataset. While centralized deep learning approaches achieve slightly higher accuracy (eg, 95.2% with light gradient boosting machine, 96.1% with convolutional neural networks), these models require full data centralization, sacrificing privacy and increasing system vulnerability.

XSL•FO
RenderX

Table . Comparative performance on the PIMA Indians Diabetes Dataset.

Model or study	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)	Notes
FedEnTrust	84.2	84.6	88.6	86.4	Federated ensemble with adaptive weighted voting and blockchain smart contract integration
ML[a] classifiers approach [28]	95.2	N/A[b]	N/A	N/A	Centralized; evaluated multiple classifiers (LR[c], XGB[d], GB[e], DT[f], ET[g], RF[h], and LGBM[i]) on PIMA Indians dataset; best accuracy achieved by LGBM
Recursive feature elimination with a gated recurrent unit RFE-GRU[j] [29]	90.7	90.5	90.7	90.5	Centralized; utilized RFE-GRU on PIMA Dataset
Hybrid classification approach [30]	83.1	N/A	64.8	N/A	Centralized; applied SVM[k], RF, DT, naive Bayes with K-means preprocessing; best accuracy achieved by SVM
Three predictive algorithms [31]	77.1	N/A	N/A	N/A	Centralized; applied LR, RF, and ANN[l]; LR achieved the best accuracy (77.10%) with AUC[m] 0.83 over RF and ANN
Soft voting ensemble [32]	79.1	73.1	71.6	80.9	Centralized; combined RF, LR, and naive Bayes classifiers
Ensemble hierarchical model [33]	83.1	25.0 (positive)/98.6 (negative)	38.4 (positive)/90.2 (negative)	82.8	Centralized; applied DT and LR, fused by neural network
Stacking ensemble [25]	77.1	N/A	N/A	N/A	Centralized; stacking ensemble of ML models; accuracy achieved using cross-validation protocol
Deep learning pipeline [34]	92.3	N/A	N/A	N/A	Centralized; deep learning pipeline using VAE[n] for data augmentation, SAE[o] for feature augmentation, and CNN[p] for classification
Deep CNN with correlation-based features [35]	96.1	94.4	94.4	94.5	Centralized; applied deep CNN with feature selection based on correlation
Blockchain-FL with adaptive DP [20]	82.7	N/A	N/A	N/A	Decentralized; implemented federated learning with differential privacy using blockchain technology

[a]ML: machine learning.

[b]N/A: not applicable.

XSL•FO
RenderX

[c]LR: logistic regression.
[d]XGB: extreme gradient boosting.
[e]GB: gradient boosting.
[f]DT: decision tree.
[g]ET: extra tree.
[h]RF: random forest.
[i]LGBM: light gradient boosting machine.
[j]RFE-GRU: Recursive Feature Elimination with Gated Recurrent Unit.
[k]SVM: support vector machine.
[l]ANN: artificial neural network.
[m]AUC: area under the curve.
[n]VAE: variational autoencoder.
[o]SAE: stacked autoencoder.
[p]CNN: convolutional neural network.

In contrast, FedEnTrust improves over recent decentralized models, such as blockchain-integrated FL with differential privacy (accuracy≈82.7%), by incorporating ensemble learning and adaptive aggregation. Despite the constraints of data fragmentation and heterogeneity, our framework maintains robust performance across all key metrics, including precision (84.6%), recall (88.6%), and F_1-score (86.4%).

FedEnTrust achieves a favorable trade-off between privacy, generalizability, and computational practicality, making it well suited for real-world deployment in regulated health care environments.

Ethical AI Considerations: Fairness, Transparency, and Accountability

Ethical Framework

Ethical concerns are central to the deployment of AI systems in health care, where unequal access to computational resources and imbalanced data distributions may inadvertently create or reinforce model biases. FedEnTrust incorporates several design principles aligned with emerging ethical AI guidelines, including those recommended by the World Health Organization and major AI governance frameworks.

Fairness Across Heterogeneous Participants

Health care institutions vary substantially in data volume, demographic composition, and computational capacity, which can introduce systematic bias in collaborative learning systems. FedEnTrust is designed to mitigate such bias by supporting heterogeneity-aware participation, allowing low-resource nodes to contribute using models aligned with their capabilities without sacrificing predictive performance. Adaptive weight clipping is applied during aggregation to prevent high-resource institutions from disproportionately dominating the global ensemble. In addition, temperature-calibrated soft labels are used to reduce overconfidence from models trained on larger or more homogeneous datasets, while confidence thresholding ensures that noisy or low-confidence predictions are not propagated across participants. Together, these mechanisms promote more balanced influence across diverse health care contributors and support fairer model outcomes in heterogeneous federated environments.

Transparency and Auditability

Transparency in FedEnTrust is enabled through the blockchain-based coordination layer, which provides immutable audit trails for all update submissions and verifiable records of role validation events. Each model contribution is traceably logged, allowing the system to record which institutions participated in and influenced each training round. This tamper-resistant logging mechanism enhances accountability, supports post hoc auditing, and increases trust among participating health care entities without exposing sensitive data or model parameters.

Privacy and Data Minimization

FedEnTrust adheres to privacy-by-design principles:

- Raw patient data remain strictly on the device
- Only soft-label vectors and hashed metadata are transmitted
- No identifiable information is stored on-chain, supporting HIPAA, GDPR, and similar regulatory frameworks

Role-based access ensures that only authorized clinical entities may participate.

Accountability and Governance

The multivalidator consensus layer enables shared governance rather than reliance on a single coordinating institution. This creates a more accountable decision-making process and aligns with ethical expectations for distributed medical AI systems.

Blockchain Performance and Practical Considerations

Implementation Considerations

Beyond empirical accuracy and security validation, the practical deployment of blockchain-enabled FL systems in health care requires careful consideration of scalability, cost, and regulatory compliance. While the blockchain layer in FedEnTrust demonstrated stable and reliable performance under controlled experimental conditions, real-world health care environments introduce additional operational and governance challenges. This section discusses key practical considerations and outlines how FedEnTrust is designed to address them.

Scalability and Throughput

Public blockchain platforms, such as the Ethereum main net, face inherent constraints related to transaction throughput, block

XSL•FO
RenderX

confirmation latency, and network congestion. These limitations can lead to unpredictable delays and may not support the repeated coordination required across multiple FL rounds. To address this, FedEnTrust is designed for deployment on private or consortium-based Ethereum networks, where consensus parameters, block times, and validator participation can be tailored to health care workflows. Such configurations enable deterministic execution and consistent performance, as observed in our evaluation. Nevertheless, large-scale deployments involving many institutions may require additional enhancements, including optimized validator load balancing, hierarchical or sharded blockchain structures, and integration with layer-2 scaling mechanisms to further increase throughput.

Cost Variability and Resource Requirements

In public blockchain environments, gas fees fluctuate dramatically based on network conditions, resulting in variable operational costs for smart contract execution. This variability is incompatible with cost-sensitive health care environments. Deploying FedEnTrust on a private Ethereum network eliminates transaction fees and allows institutions to control computational and storage overhead. However, operating such networks requires institutional commitment to maintain validator nodes, ensure uptime, and manage governance policies. Future work will investigate cost-benefit trade-offs between private, hybrid, and layer-2 blockchain configurations for FL.

Regulatory and Compliance Constraints

Health care systems must comply with strict privacy regulations such as HIPAA, GDPR, and provincial or national data-protection laws. These frameworks introduce challenges, such as prohibiting the storage of patient data or identifiers on-chain, requiring transparent audit trails for collaborative analytics, and ensuring that cross-institution coordination adheres to data-sharing agreements.

FedEnTrust addresses these concerns by storing only hashed metadata and role-verification entries on-chain, keeping soft labels and model outputs entirely off-chain. However, real-world deployment requires integration with institutional governance mechanisms to ensure compliance documentation, legal interoperability among institutions, and formal auditing procedures.

Generalizability to Multimodal and Longitudinal Health Care Data

Although the PIMA dataset provides a controlled benchmark for evaluating prediction accuracy, it does not reflect the complexity of real-world clinical environments. Modern health care systems generate multimodal data that may include structured electronic health record fields, laboratory values, medical imaging, clinician notes, and continuous wearable sensor streams. Additionally, many health conditions, including diabetes, require longitudinal modeling to capture evolving physiological states over time.

FedEnTrust is designed to naturally extend to these scenarios. The framework's heterogeneity-aware model assignment allows each participant to select model architectures aligned with its data modality and computational resources. For example, hospitals could train sequence models (eg, long short-term memories or transformers) on longitudinal EHR data, while wearable devices may contribute short-term physiological features via lightweight SVM or tree-based models. The knowledge-distillation component operates on probability distributions and is therefore agnostic to model type, enabling soft-label fusion across diverse modalities and temporal structures. This capability is particularly suitable for integrating outputs from time-series models, tabular models, and sensor analytics.

The blockchain-based coordination layer also supports generalization, as its role-based validation and update logging apply to any model output regardless of modality. Future work will apply FedEnTrust to multicenter datasets such as MIMIC-IV, NHANES, and integrated wearable–EHR cohorts to evaluate its performance under more heterogeneous and clinically realistic conditions.

Limitations

Despite promising results, several limitations remain:

- Dataset representativeness: The PIMA dataset is limited in scope and population diversity. Future work should evaluate FedEnTrust on broader, real-world datasets from varied demographics and geographies.
- Extreme client heterogeneity: Devices with ultra-low resources may still face difficulties in real-time model adaptation. Exploring ultra-lightweight architectures and communication compression techniques is a key next step.
- Controlled blockchain simulation: Our blockchain operations were simulated under stable conditions. Future deployment on public testnets or mainnets is necessary to assess real-world transaction delays, scalability, and cost variability.
- Advanced threat modeling: While the smart contract blocks unauthorized actions, adversarial behaviors such as collusion or model poisoning were not addressed. Future extensions may integrate anomaly detection and audit trails to enhance system resilience.

Although the PIMA Indians Diabetes Dataset is a well-established benchmark for evaluating diabetes prediction models, its limited demographic diversity and relatively small size restrict the generalizability of the findings. The simulated heterogeneous environment in Table 2, while constructed to reflect realistic participant variability, does not fully replicate the complexity of multi-institution health care settings, where differences in clinical practice, sensor characteristics, and patient demographics lead to substantially wider non-IID distributions. Accordingly, the results presented here should be viewed as a controlled feasibility demonstration rather than a comprehensive real-world validation.

Conclusions

This study presents FedEnTrust, a secure and intelligent federated ensemble learning framework for privacy-preserving diabetes prediction. Our approach addresses key challenges in decentralized health care AI, including data privacy, system trust, and participant heterogeneity, without requiring access to raw patient data.

XSL•FO
RenderX

By integrating knowledge distillation and adaptive ensemble aggregation, the framework enables resource-aware contributions from a diverse range of participants, from high-performance hospital systems to low-power personal devices. The experimental results demonstrate consistent improvements in predictive performance across all participants, validating both the effectiveness and inclusiveness of the design.

A central innovation is the blockchain-enabled coordination layer, which ensures secure registration, role-based access control, and verifiable model updates. Smart contract simulations confirm the system's efficiency, low latency, and robustness against unauthorized actions, supporting scalable and tamper-resistant deployment in health care environments.

In sum, FedEnTrust offers a practical, scalable solution for secure, decentralized medical AI, balancing privacy, performance, and trust. Future work will extend this framework to additional clinical domains, multisite studies, and dynamic personalization for broader impact in real-world health care.

Funding

The authors express their gratitude to the funding provided to support this study from the National Science Foundation with award number 2218046.

Conflicts of Interest

None declared.

References

1. Ong KL, Stafford LK, McLaughlin SA. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
2. Jaiswal V, Negi A, Pal T. A review on current advances in machine learning based diabetes prediction. Prim Care Diabetes 2021 Jun;15(3):435-443. [doi: 10.1016/j.pcd.2021.02.005] [Medline: 33642253]
3. Nomura A, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial intelligence in current diabetes management and prediction. Curr Diab Rep 2021 Dec 13;21(12):61. [doi: 10.1007/s11892-021-01423-2] [Medline: 34902070]
4. Ahmed U, Issa GF, Khan MA, et al. Prediction of diabetes empowered with fused machine learning. IEEE Access 2022;10:8529-8538. [doi: 10.1109/ACCESS.2022.3142097]
5. Choudhur O, Gkoulalas-Divanis A, Salonidis T, et al. Differential privacy-enabled federated learning for sensitive health data. arXiv. Preprint posted online on Oct 7, 2019. [doi: 10.48550/arXiv.1910.02578]
6. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. J Healthc Inform Res 2021;5(1):1-19. [doi: 10.1007/s41666-020-00082-4] [Medline: 33204939]
7. Zisis K, Pavi E, Geitona M, Athanasakis K. Real-world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process. J Pharm Pharm Sci 2024;27:12302. [doi: 10.3389/jpps.2024.12302] [Medline: 38481726]
8. Kairou P, McMahan HB, Avent B, et al. Advances and open problems in federated learning. arXiv. Preprint posted online on Dec 10, 2019. [doi: 10.48550/arXiv.1912.04977]
9. Islam H, Mosa A. A federated mining approach on predicting diabetes-related complications: demonstration using real-world clinical data. AMIA Annu Symp Proc 2021;2021:556-564. [Medline: 35308968]
10. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare (Basel) 2024 Dec 22;12(24):2587. [doi: 10.3390/healthcare12242587] [Medline: 39766014]
11. Liu J, Chen C, Li Y, et al. Enhancing trust and privacy in distributed networks: a comprehensive survey on blockchain-based federated learning. Knowl Inf Syst 2024 Aug;66(8):4377-4403. [doi: 10.1007/s10115-024-02117-3]
12. Lu Z, Pan H, Dai Y, Si X, Zhang Y. Federated learning with non-IID data: a survey. IEEE Internet Things J 2024 Jun;11(11):19188-19209. [doi: 10.1109/JIOT.2024.3376548]
13. Yang G, Edwards B, Bakas S, et al. Federated learning as a catalyst for digital healthcare innovations. Patterns (N Y) 2024 Jul 12;5(7):101026. [doi: 10.1016/j.patter.2024.101026] [Medline: 39081574]
14. Joshi M, Pal A, Sankarasubbu M. Federated learning for healthcare domain—pipeline, applications and challenges. ACM Trans Comput Healthcare 2022 Oct 31;3(4):1-36. [doi: 10.1145/3533708]
15. Hasan MR, Li Q, Saha U, Li J. Decentralized and secure collaborative framework for personalized diabetes prediction. Biomedicines 2024 Aug 21;12(8):1916. [doi: 10.3390/biomedicines12081916] [Medline: 39200380]
16. Patel VA, Bhattacharya P, Tanwar S, et al. Adoption of federated learning for healthcare informatics: emerging applications and future directions. IEEE Access 2022;10:90792-90826. [doi: 10.1109/ACCESS.2022.3201876]
17. Wulfert L, Wiede C, Grabmaier A. TinyFL: on-device training, communication and aggregation on a microcontroller for federated learning. Presented at: 2023 21st IEEE Interregional NEWCAS Conference (NEWCAS); Jun 26-28, 2023; Edinburgh, United Kingdom p. 1-5. [doi: 10.1109/NEWCAS57931.2023.10198040]

XSL•FO
RenderX

18. Niknam S, Dhillon HS, Reed JH. Federated learning for wireless communications: motivation, opportunities, and challenges. IEEE Commun Mag 2020 May;58(6):46-51. [doi: 10.1109/MCOM.001.1900461]
19. Shayan M, Fung C, Yoon CJM, Beschastnikh I. Biscotti: a blockchain system for private and secure federated learning. IEEE Trans Parallel Distrib Syst 2021 Jul;32(7):1513-1525. [doi: 10.1109/TPDS.2020.3044223]
20. Chang Y, Fang C, Sun W. A blockchain-based federated learning method for smart healthcare. Comput Intell Neurosci 2021;2021:4376418. [doi: 10.1155/2021/4376418] [Medline: 34868289]
21. Kang J, Xiong Z, Niyato D, Zou Y, Zhang Y, Guizani M. Reliable federated learning for mobile networks. IEEE Wireless Commun 2020 Apr;27(2):72-80. [doi: 10.1109/MWC.001.1900119]
22. Zhang H, Li G, Zhang Y, Gai K, Qiu M. Blockchain-based privacy-preserving medical data sharing scheme using federated learning. In: Knowledge Science, Engineering and Management: Springer; 2021:634-646. [doi: 10.1007/978-3-030-82153-1_52]
23. Tripathi G, Ahad MA, Casalino G. A comprehensive review of blockchain technology: underlying principles and historical background with future challenges. Decis Anal J 2023 Dec;9:100344. [doi: 10.1016/j.dajour.2023.100344]
24. Shalan M, Hasan MR, Bai Y, Li J. Enhancing smart home security: blockchain-enabled federated learning with knowledge distillation for intrusion detection. Smart Cities 2025 Feb;8(1):35. [doi: 10.3390/smartcities8010035]
25. Reza MS, Amin R, Yasmin R, Kulsum W, Ruhi S. Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. Heliyon 2024 Jan 30;10(2):e24536. [doi: 10.1016/j.heliyon.2024.e24536] [Medline: 38312584]
26. Chang V, Bailey J, Xu QA, Sun Z. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Comput Appl 2022 Mar 24:1-17. [doi: 10.1007/s00521-022-07049-z] [Medline: 35345556]
27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2011;16:321-357. [doi: 10.1613/jair.953]
28. Ahamed BS, Arya MS, Nancy VAO. Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques. Front Comput Sci 2022 May;4. [doi: 10.3389/fcomp.2022.835242]
29. Shams MY, Tarek Z, Elshewey AM. A novel RFE-GRU model for diabetes classification using PIMA Indian dataset. Sci Rep 2025 Jan 6;15(1):982. [doi: 10.1038/s41598-024-82420-9] [Medline: 39762262]
30. Edeh MO, Khalaf OI, Tavera CA, et al. A classification algorithm-based hybrid diabetes prediction model. Front Public Health 2022;10:829519. [doi: 10.3389/fpubh.2022.829519] [Medline: 35433625]
31. Darolia A, Chhillar RS. Analyzing three predictive algorithms for diabetes mellitus against the PIMA Indians Dataset. ECS Trans 2022 Apr 24;107(1):2697-2704. [doi: 10.1149/10701.2697ecst]
32. Kumari S, Kumar D, Mittal M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int J Cogn Comput Eng 2021 Jun;2:40-46. [doi: 10.1016/j.ijcce.2021.01.001]
33. Abedini M, Bijari A, Banirostam T. Classification of Pima Indian Diabetes Dataset using ensemble of decision tree, logistic regression and neural network. Int J Adv Res Comput Commun Eng 2020 Jul 30;9(7):1-4. [doi: 10.17148/IJARCCE.2020.9701]
34. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. Comput Methods Programs Biomed 2021 Apr;202:105968. [doi: 10.1016/j.cmpb.2021.105968] [Medline: 33631638]
35. Thushara A. An efficient Alzheimer's disease prediction based on MEPC-SSC segmentation and momentum geo-transient MLPs. Comput Biol Med 2022 Dec;151(Pt A):106247. [doi: 10.1016/j.compbiomed.2022.106247] [Medline: 36375415]

Abbreviations

AI: artificial intelligence
FL: federated learning
GDPR: General Data Protection Regulation
HIPPA: Health Insurance Portability and Accountability Act
IID: independent and identically distributed
KNN: k-nearest neighbors
RBAC: role-based access control
SVC: support vector classifier
SVM: support vector machine

XSL•FO
RenderX

Edited by S Li; submitted 16.Jun.2025; peer-reviewed by M Gheisari, S Patel; revised version received 13.Dec.2025; accepted 17.Dec.2025; published 26.Jan.2026.

Please cite as:
Hasan MR, Li J
Privacy-Preserving Collaborative Diabetes Prediction in Heterogeneous Health Care Systems: Algorithm Development and Validation of a Secure Federated Ensemble Framework
JMIR Diabetes 2026;11:e79166
URL: https://diabetes.jmir.org/2026/1/e79166
doi:10.2196/79166

© Md Rakibul Hasan, Juan Li. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 26.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Toward Robust AI-Assisted Dietary Assessment for Diabetes Self-Management: Quantifying and Decomposing Large Language Model Prediction Variability From Meal Images

Zhaohua Wang[1], MS; Daniel Lane[1], MS, MBA; Kayo Waki[1,2], MD, MPH, PhD

[1]Department of Health and Behavioral Informatics & Therapeutics (HABIT), Graduate School of Medicine, The University of Tokyo, Faculty of Medicine Bldg. 2, 2nd Floor, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

[2]Department of Diabetes and Metabolic Diseases, The University of Tokyo Hospital, Tokyo, Japan

Corresponding Author:
Kayo Waki, MD, MPH, PhD
Department of Health and Behavioral Informatics & Therapeutics (HABIT), Graduate School of Medicine, The University of Tokyo, Faculty of Medicine Bldg. 2, 2nd Floor, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Nutrient estimation from meal images by multimodal large language models can support diabetes self-management, but its robustness is limited by variability driven by both sensitivity to visual presentation and inherent model instability. Accounting for and mitigating this variability is essential for robust AI-assisted dietary assessment.

(JMIR Diabetes 2026;11:e102715) doi:10.2196/102715

KEYWORDS

multimodal large language models; artificial intelligence; prediction variability; dietary assessment; nutrient estimation

Introduction

Dietary assessment is fundamental to diabetes self-management. AI has advanced dietary assessment [1]. In particular, multimodal large language models (LLMs) such as the GPT family offer an automated and rapid alternative to manual approaches for estimating nutrient content from meal images [2]. The accuracy of LLMs has been validated [3], but the precision remains insufficiently investigated. Because LLMs are inherently conditional and probabilistic, their predictions inevitably exhibit variability. Low precision can reduce the accuracy of nutrient estimation and reduce patient confidence in AI-assisted self-management tools. Characterizing and addressing this variability is essential for achieving robust AI-assisted dietary assessment for diabetes self-management.

In practice, a meal can be photographed from arbitrary camera viewpoints or with arbitrary dish arrangements. Such visual variations can lead to different predictions for different images of the same meal, resulting in between-image variability. At the same time, even under deterministic settings, LLMs can still exhibit residual stochasticity [4]. Such random processes can lead to different predictions for the same image of the same meal, resulting in within-image variability. Between-image variability reflects the model's sensitivity to visual presentation, and within-image variability reflects its internal stability. Together, they constitute the prediction variability for the same meal, reflecting the precision of LLMs.

Based on paired images with different camera viewpoints or dish arrangements, we quantified the prediction variability of GPT-4o and GPT-5.1 and decomposed it into between-image variability and within-image variability, with the aim of characterizing the precision of LLM-based nutrient estimation from meal images.

Methods

We collected 20 meal-image pairs, each pair depicting the same meal: 10 with different camera viewpoints and 10 with different dish arrangements. We evaluated 2 multimodal LLMs, GPT-4o and GPT-5.1. For each image, we issued 60 independent single-turn requests to estimate the meal's energy, carbohydrates, protein, fat, fiber, and salt content with fixed prompt and deterministic parameter settings (Multimedia Appendix 1). The prompt specified neither an image-parsing process nor a nutritional reference, only noting that the meal was eaten in Japan.

Let $Y=\mu+\varepsilon$ denote the nutrition content prediction, which can be decomposed into the expected prediction μ and independent stochastic noise ε with $E(\varepsilon)=0$. For an image pair (A,B), we define the prediction difference as $D_{A,B}=Y_A-Y_B$. The expected prediction variability on the linear scale is E. Since each request is independent, we can assume the covariance of the stochastic noises $Cov(\varepsilon_A,\varepsilon_B)=0$. Consequently, we can additively decompose the expected prediction variability on a squared scale into between-image variability and within-image variability: $E[D_{A,B}2]=(\mu_A-\mu_B)2+Var(Y_A)+Var(Y_B)$.

XSL•FO
RenderX

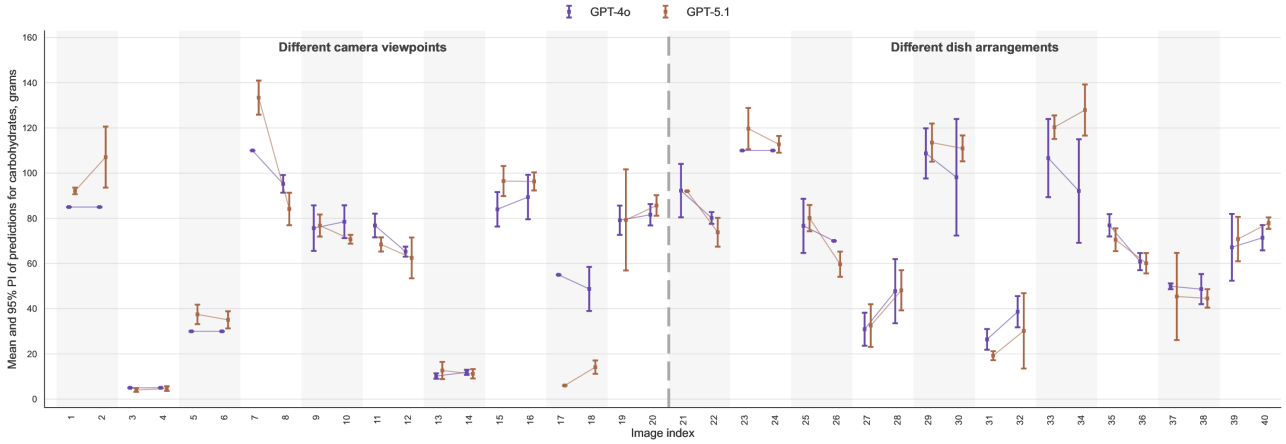
In this decomposition, E[DA,B2] represents the prediction variability and reflects precision, (μA−μB)2 represents the between-image variability and reflects visual sensitivity, and Var(YA)+Var(YB) represents the within-image variability and reflects internal stability. To ensure interpretability, we took the square root of each variability component to convert it into a root-mean-square (RMS) magnitude on the linear scale.

For each measure, we normalized it by the pooled mean prediction of the meal (μA+μB)2, to obtain the dimensionless relative magnitudes. We then derived global results by aggregating pair-level local results using bootstrap resampling, treating each image pair as a sampling unit with 10,000 iterations. Nutrient-level estimates were obtained by averaging each measure across all image pairs. Model-level estimates were obtained by averaging the pair-level mean of each measure (averaged across the 6 nutrients) across all image pairs.

Results

Qualitative inspection revealed variability in predictions for the same meal. The dispersion of repeated predictions within one individual image evidenced within-image variability, and the difference of mean predictions between 2 paired images evidenced between-image variability (Figure 1 and Multimedia Appendix 2).

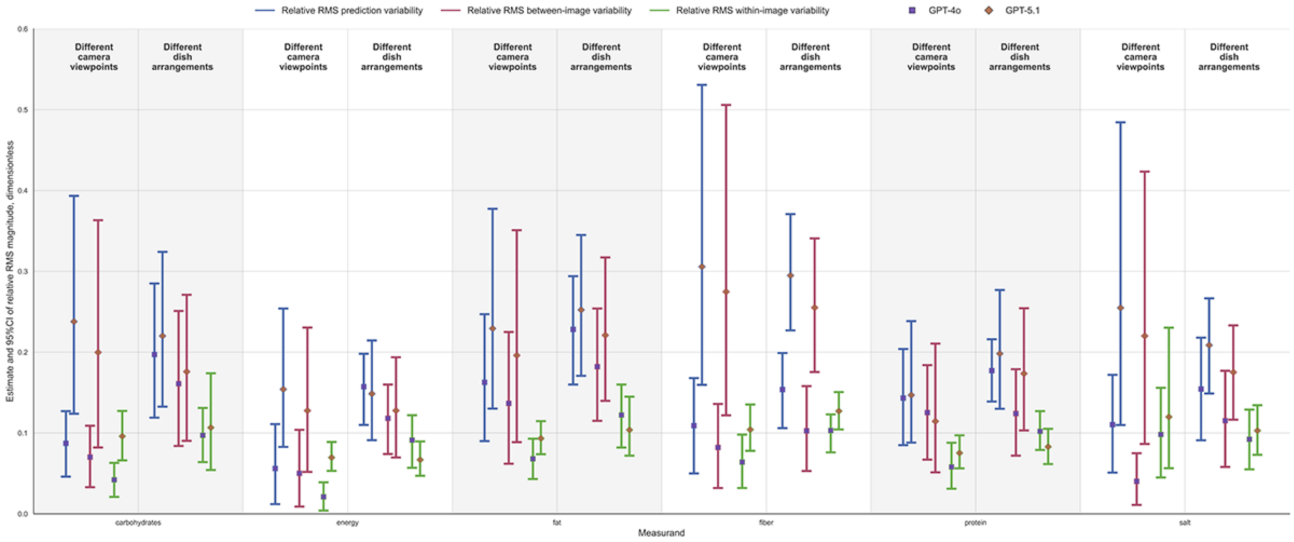
Figure 1. Prediction distributions for carbohydrate content of each image on GPT-4o and GPT-5.1. The plot shows the mean prediction and 95% prediction interval (PI) for carbohydrates content of each image on GPT-4o and GPT-5.1, derived from 60 repeated runs per image. Connected points indicate paired images depicting the exact same physical meal, with different visual presentation: different camera viewpoints (image pair 1 vs 2 to image pair 19 vs 20) or different dish arrangements (image pair 21 vs 22 to image pair 39 vs 40). The vertical distance between these points reflects the between-image variability, while the length of the error bars reflects the within-image variability.



Mean absolute magnitudes directly quantified the prediction variability in natural units (Multimedia Appendix 3). Variance decomposition further quantified and decomposed the prediction variability. Regarding the measurands, both GPT-4o and GPT-5.1 showed smaller relative RMS magnitudes for energy than for the other nutrients. Among those nutrients, macronutrients such as carbohydrates and protein exhibited comparatively lower magnitudes (Figure 2). Regarding the response to image variations, GPT-4o showed higher sensitivity to dish arrangement changes than to camera viewpoint changes. In contrast, GPT-5.1 showed similarly high sensitivity to both image variations (Multimedia Appendix 4). Regarding the models, GPT-5.1 showed greater overall variability than GPT-4o. It had a higher relative RMS prediction variability, which was driven primarily by larger between-image variability, whereas its within-image variability was similar to that of GPT-4o (Multimedia Appendix 4).

XSL•FO
RenderX

Figure 2. Global estimates of prediction variability and its decomposition for each measurand on GPT-4o and GPT-5.1. The plot shows the bootstrapped aggregated estimate and 95% CI of relative root-mean-square (RMS) prediction variability, relative RMS between-image variability, and relative RMS within-image variability, derived by aggregating the relative RMS magnitudes across image pairs depicting the exact same physical meal with different camera viewpoints and image pairs with different dish arrangements, respectively.



Discussion

This study qualitatively demonstrated and quantitatively assessed the prediction variability of LLMs in estimating nutrient content from meal images. There is nonnegligible prediction variability for images of the same meal, and it can be decomposed into between-image variability from visual variation of the images and within-image variability from the inherent stochasticity of the model.

Between-image variability reflects a model's sensitivity to visual presentation. While such sensitivity is essential for capturing subtle food features, it becomes detrimental when the model overreacts to nonessential variations, such as camera viewpoints or dish arrangements. On GPT-4o, the between-image variability from dish arrangement changes was greater than that from camera viewpoint changes, as the visual variation of dish arrangement changes is often larger. On GPT-5.1, the between-image variability was higher than on GPT-4o despite its advanced architecture, which may be attributable to its increased sensitivity to subtle visual perturbations. Moreover, the between-image variability from camera viewpoint changes was similar to that from dish arrangement changes, which may also be attributable to its increased sensitivity making its response to simple camera viewpoint changes similar in intensity to its response to complex dish arrangement changes.

Within-image variability reflects a model's internal stability. It indicates the model's certainty about the estimation task. When estimating nutrients lacking visual cues such as salt or fiber, the model, unable to extract concrete evidence from the image, has to rely more on prior knowledge to probabilistically guess from an implicit distribution less constrained than those for energy and macronutrients, resulting in drastic fluctuations in the output for the same input.

Given the prediction variability from excessive sensitivity and internal instability, relying on a single prediction from a single image is insufficient to guarantee estimation reliability. Two pathways may mitigate the prediction variability: for a single meal, multiple photos with different camera viewpoints or dish arrangements can be captured, combining their results to limit between-image variability; for a single image, multiple requests can be issued to obtain repeated predictions, aggregating their results to mitigate within-image variability. However, these approaches should be balanced against user burden, latency and cost, and more deployment-oriented alternatives such as uncertainty-aware outputs and modular estimation pipelines require further study.

The predictions were not compared with true values, so the conclusions concern only precision, not the accuracy or validity of AI-based dietary assessment. In addition, given the inherent opacity of LLMs, the estimation process remains essentially a black box, making it difficult to address variability at the level of model processing.

Acknowledgments

We thank all the participants of the previous clinical trials for allowing their meal photographs to be used for secondary research.

Funding

This work was supported by internal laboratory funds.

XSL•FO
RenderX

Data Availability

The data is available from the corresponding author upon reasonable request for noncommercial use.

Authors' Contributions

Conceptualization: KW, DL, ZW
Data curation: ZW
Formal analysis: ZW
Methodology: ZW, DL, KW
Software: DL, ZW
Visualization: DL, ZW
Writing – original draft: ZW
Writing – review & editing: DL, KW.

Conflicts of Interest

KW (Chief Operating Officer) and DL (Chief Technology Officer) hold equity in WaShiLa Health, a start-up that is seeking to apply this technology commercially. ZW declares no conflicts of interest.

Multimedia Appendix 1
Experimental models, parameters, and prompts.
[DOCX File, 18 KB - diabetes_v11i1e102715_app1.docx]

Multimedia Appendix 2
Prediction distributions for energy, protein, fat, dietary fiber, and salt content of each image on GPT-4o and GPT-5.1.
[PNG File, 223 KB - diabetes_v11i1e102715_app2.png]

Multimedia Appendix 3
Mean absolute prediction variability for each measurand and image variation type on GPT-4o and GPT-5.1.
[DOCX File, 21 KB - diabetes_v11i1e102715_app3.docx]

Multimedia Appendix 4
Global estimates of prediction variability and its decomposition for each image variation types on GPT-4o and GPT-5.1.
[PNG File, 67 KB - diabetes_v11i1e102715_app4.png]

References

1. Evert AB, Dennison M, Gardner CD, et al. Nutrition therapy for adults with diabetes or prediabetes: a consensus report. Diabetes Care 2019 May;42(5):731-754. [doi: 10.2337/dci19-0014] [Medline: 31000505]
2. Lo FPW, Qiu J, Wang Z, et al. Dietary assessment with multimodal ChatGPT: a systematic analysis. IEEE J Biomed Health Inform 2024 Dec;28(12):7577-7587. [doi: 10.1109/JBHI.2024.3417280] [Medline: 38900623]
3. Fridolfsson J, Sjöberg E, Thiwång M, Pettersson S. Performance evaluation of 3 large language models for nutritional content estimation from food images. Curr Dev Nutr 2025 Oct;9(10):107556. [doi: 10.1016/j.cdnut.2025.107556] [Medline: 41081011]
4. Atıl B, Aykent S, Chittams A, et al. Non-determinism of “deterministic” LLM system settings in hosted environments. Presented at: 5th Workshop on Evaluation and Comparison of NLP Systems; Dec 23, 2025. [doi: 10.18653/v1/2025.eval4nlp-1.12]

Abbreviations

LLM: large language model
RMS: root-mean-square

XSL•FO
RenderX

Edited by GGR Sng, S Li; submitted 28.May.2026; peer-reviewed by K Asakura, KH Lin; revised version received 22.Jul.2026; accepted 20.Aug.2026; published 04.Sep.2026.

Please cite as:
Wang Z, Lane D, Waki K
Toward Robust AI-Assisted Dietary Assessment for Diabetes Self-Management: Quantifying and Decomposing Large Language Model Prediction Variability From Meal Images
JMIR Diabetes 2026;11:e102715
URL: https://diabetes.jmir.org/2026/1/e102715
doi:10.2196/102715

© Zhaohua Wang, Daniel Lane, Kayo Waki. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 4.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX