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

Research Letter

Viewpoints

Reviews

Tutorial

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

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

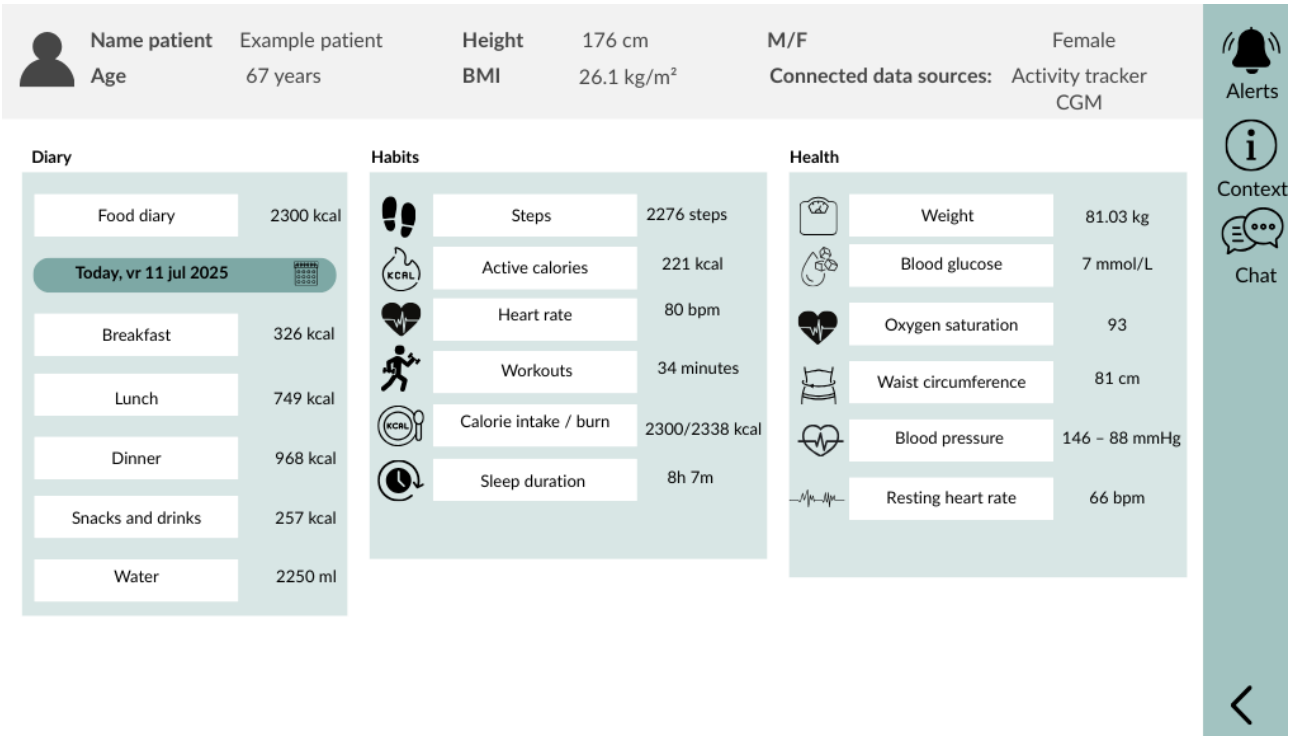
We used purposive sampling to recruit HCPs involved in T2DM care across secondary and primary care in the Twente region (Netherlands). In total, 27 HCPs were invited across both sessions. In the first focus group, 8 HCPs from secondary care, together with a lifestyle coach, were invited to participate 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.

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.

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.

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.				✓

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.		✓		

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.

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.

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.

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						

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.

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.

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.

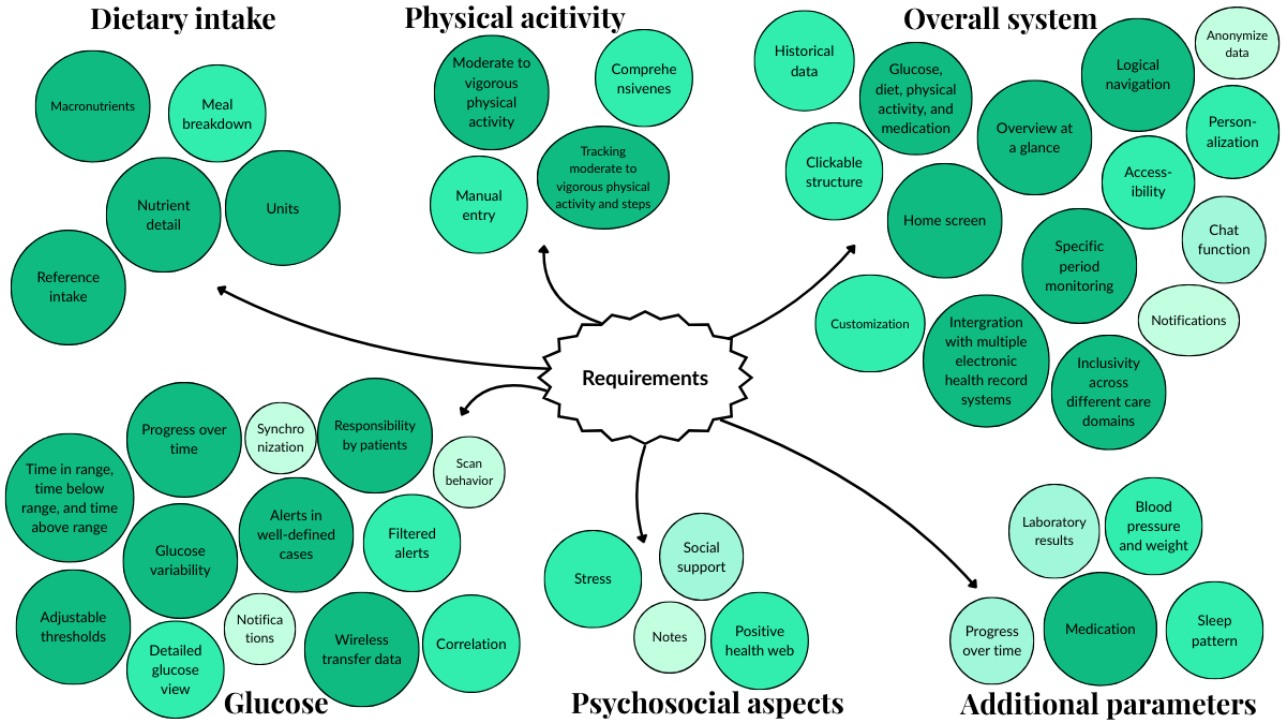
[b]HCP: health care professional.

[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.

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,

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

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

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

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]

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]

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]

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]

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

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

(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.

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):

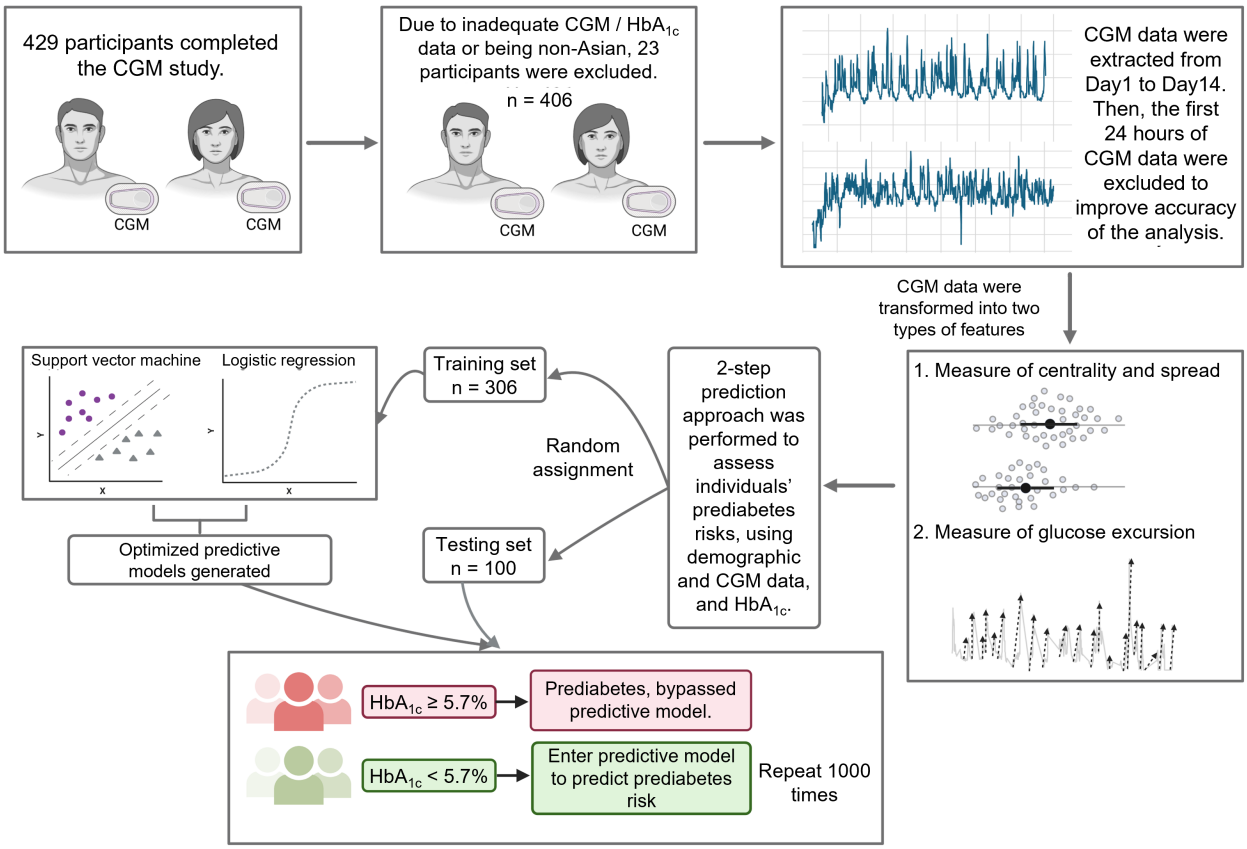
1. Demographic data encompassing age and gender, and 2 fundamental clinical measurements, namely BMI and waist-hip ratios (henceforth referred to as "Demo");
2. A combination of all the CGM summary statistics and the Demo data specified in set 1 ("CGM");
3. The CGM dataset listed in set 2, supplemented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

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).

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).

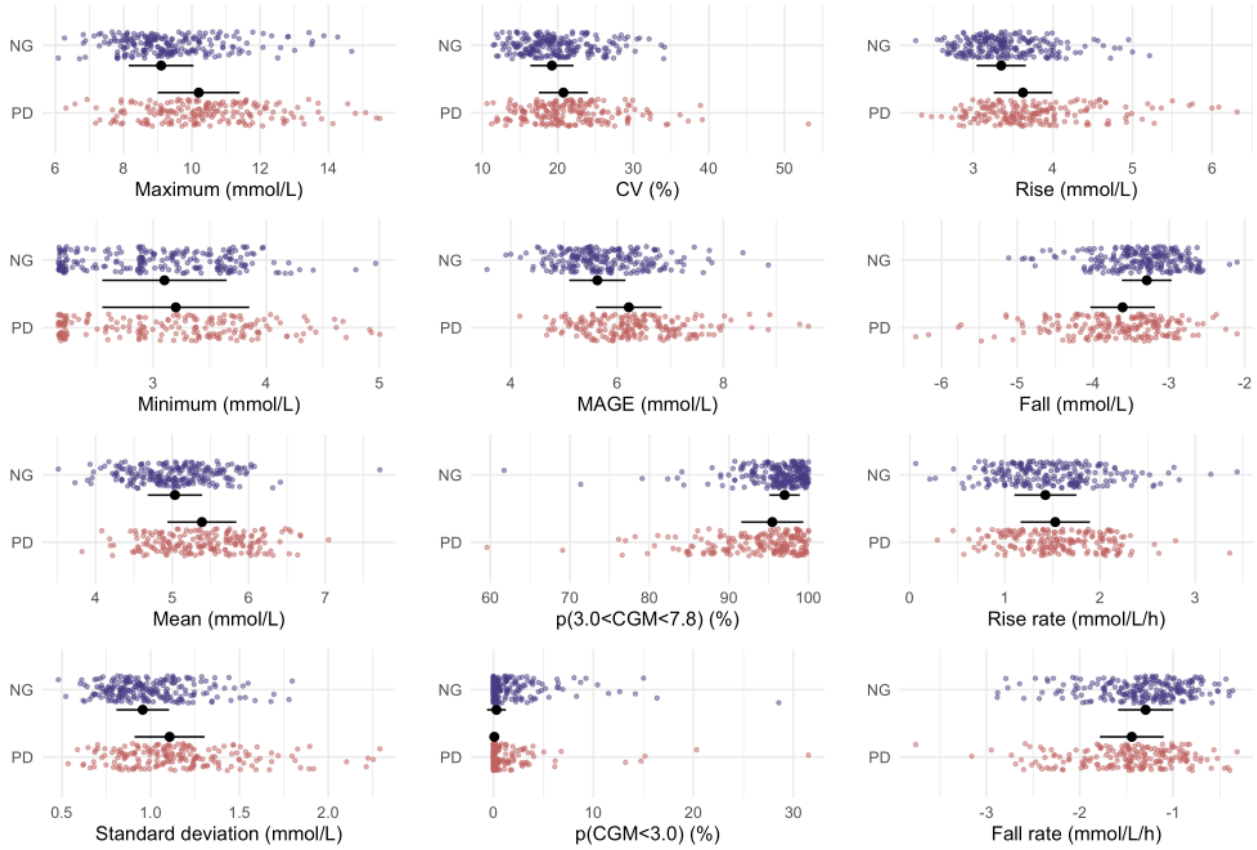
Variable	Total (N=406)	Normoglycemia (n=217)	Prediabetes (n=189)	*P* value
Sex, n (%)				.40
Male	170 (41.9)	95 (43.8)	75 (39.7)	
Female	236 (58.1)	122 (56.2)	114 (60.3)	
Ethnicity, n (%)				.46
Chinese	267 (65.8)	143 (65.9)	124 (65.6)	
Malay	57 (14.0)	27 (12.4)	30 (15.9)	
Indian	63 (15.5)	34 (15.7)	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).

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).

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.

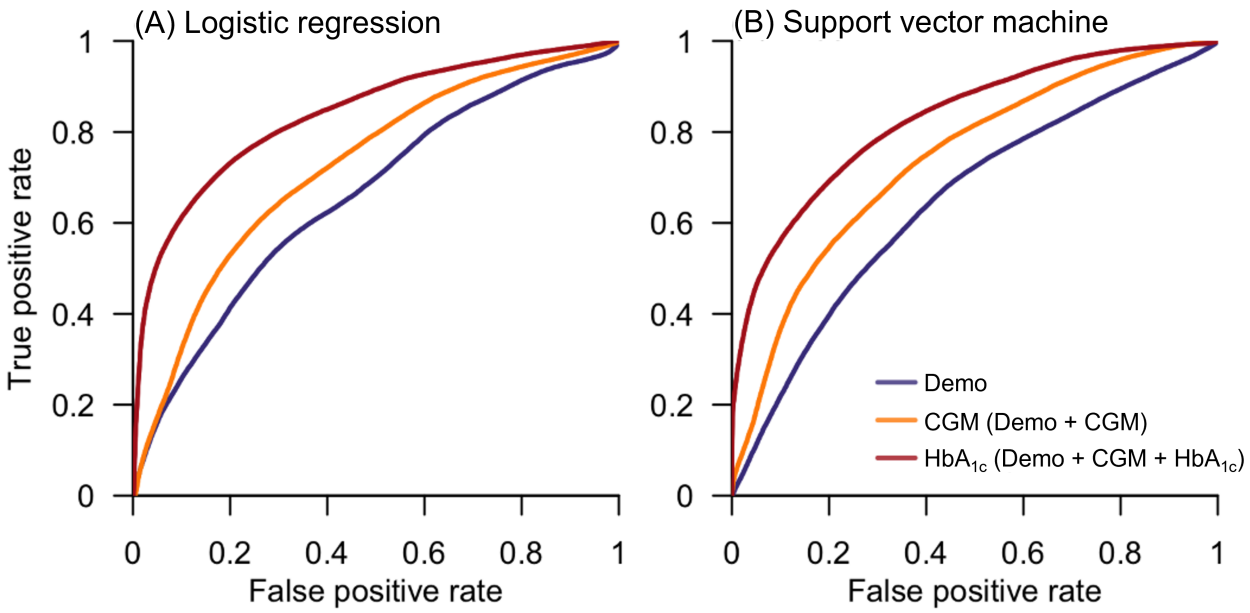


Table . Prediction accuracy of prediabetes, in terms of misclassification rates, specificity, and sensitivity, for model variants using distinct predictor sets and classification algorithms. The data are presented as the mean and 95% confidence intervals, derived from 1000 random splits into training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

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.

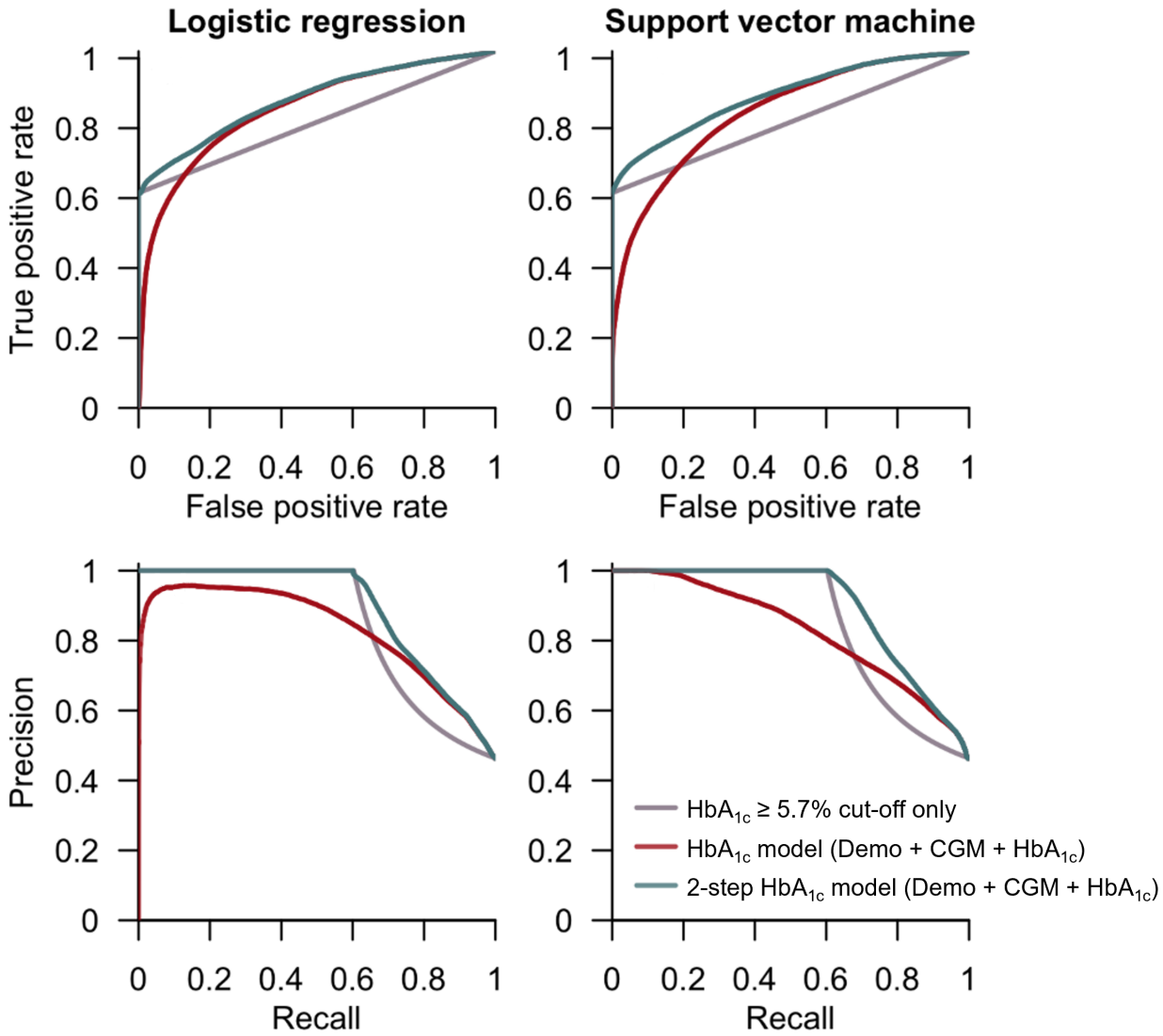
[c]area under the precision recall curve.

[d]CGM: continuous glucose monitoring.

[e]LR: logistic regression.

[f]SVM: support vector machine.

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).

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

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

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]

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

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

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

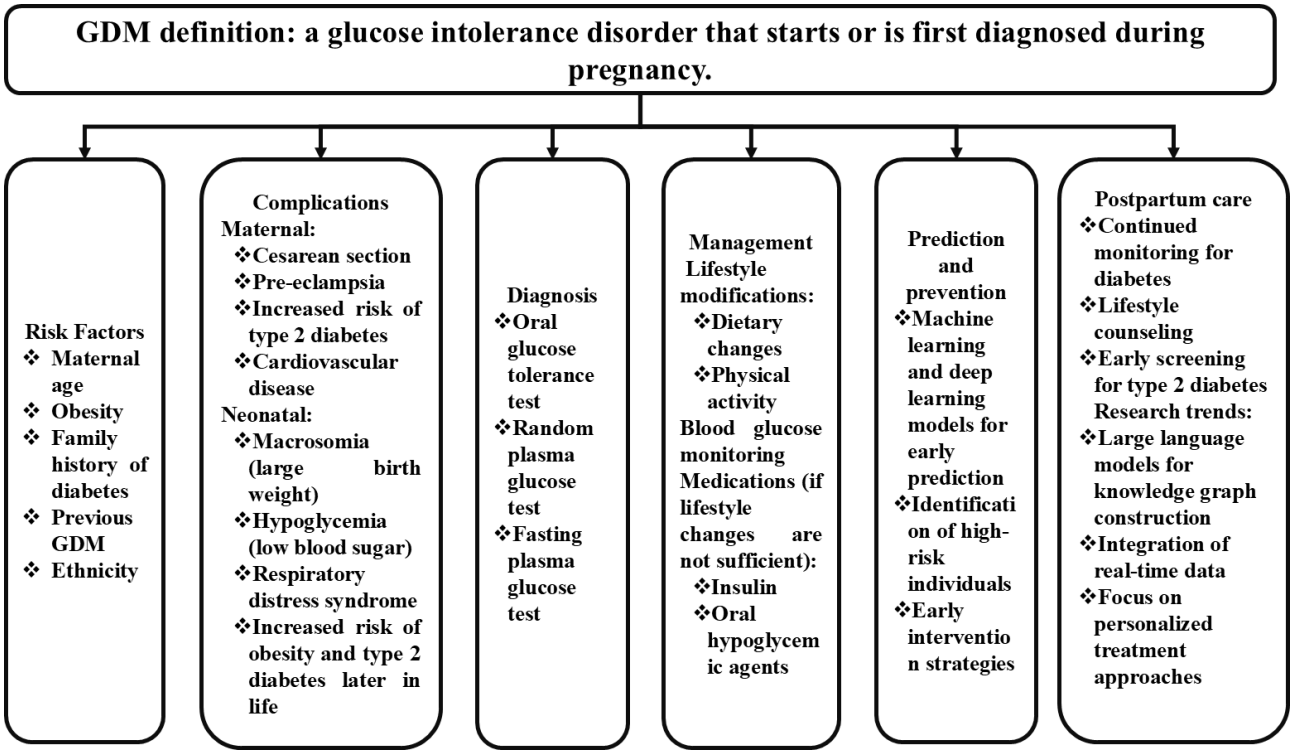
Introduction

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

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.

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.

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.

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.

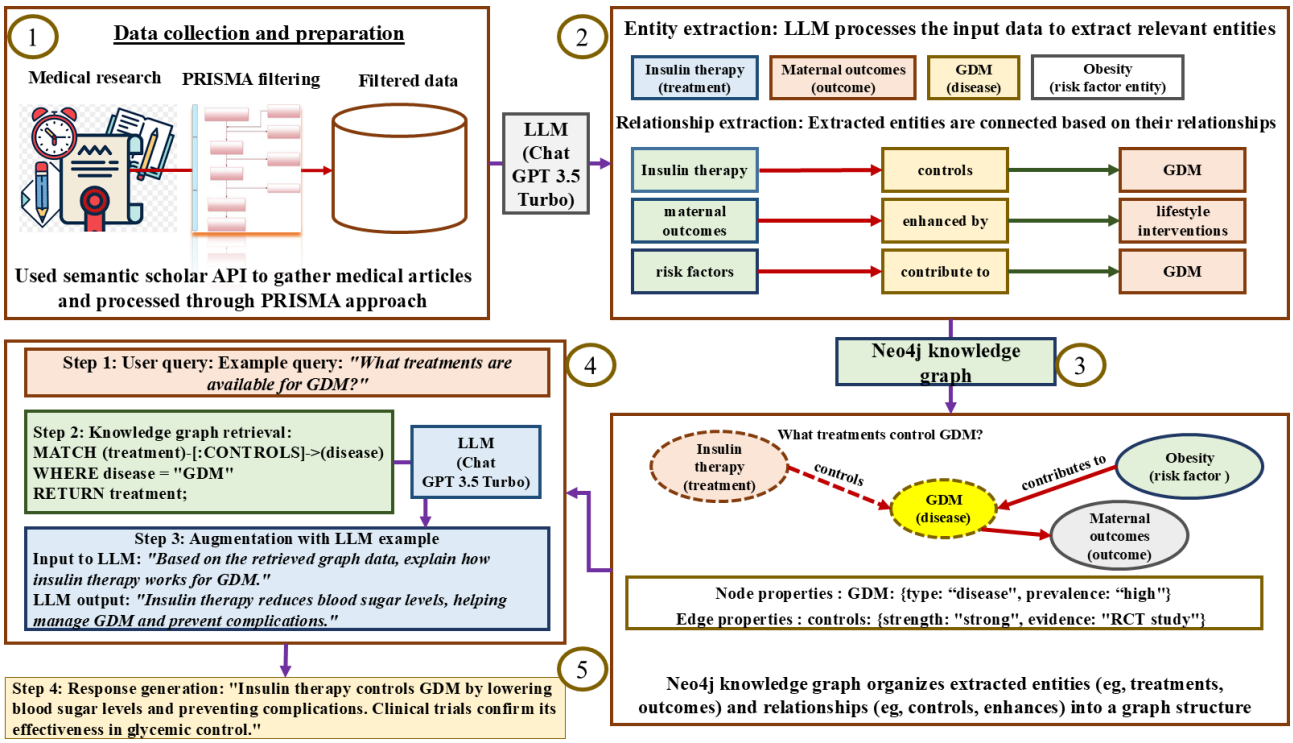
Methods

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.



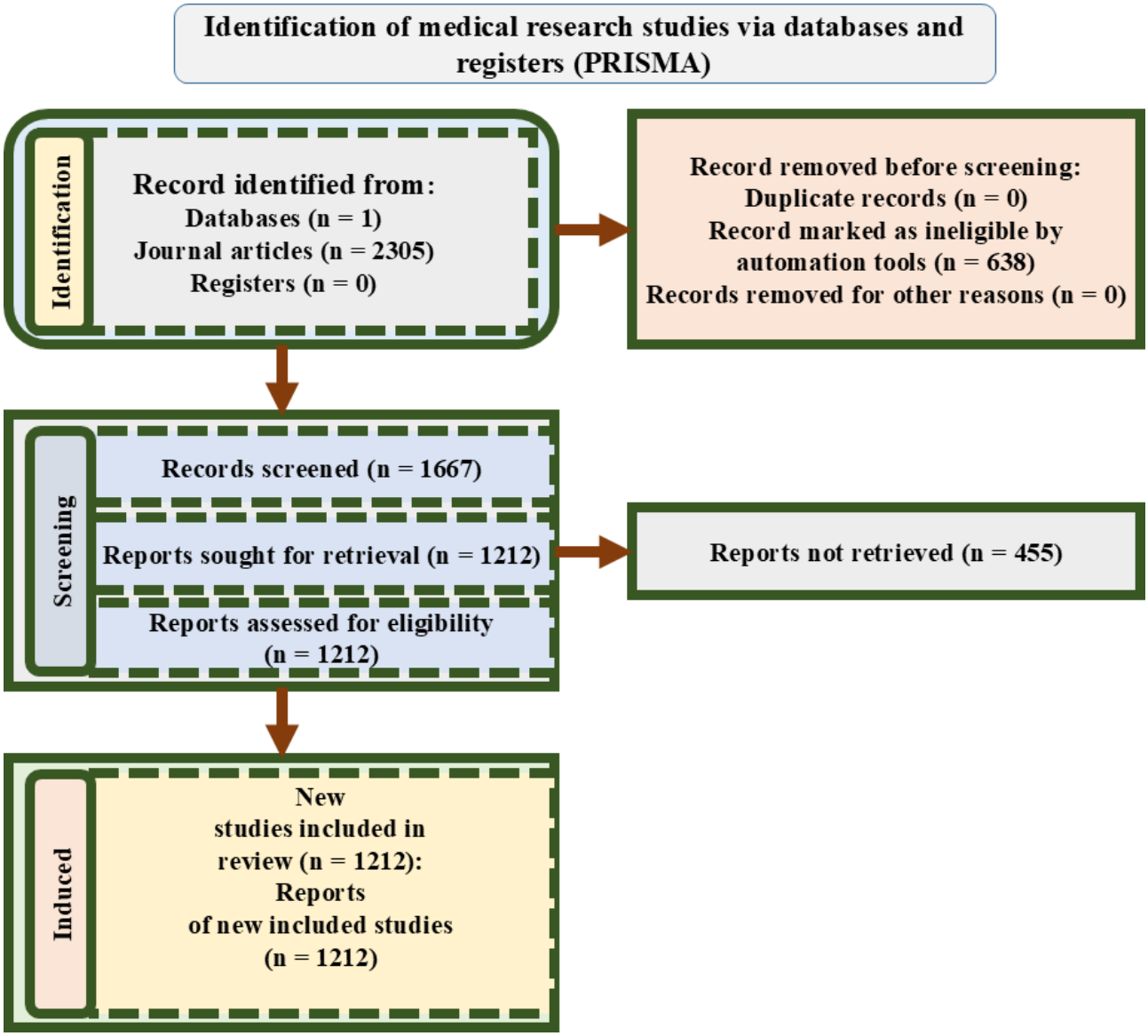
- Data collection and preparation: The Semantic Scholar API retrieved relevant research articles on GDM interventions. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)–guided filtering process was applied to ensure that inclusion criteria were met, resulting in a refined 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.

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.

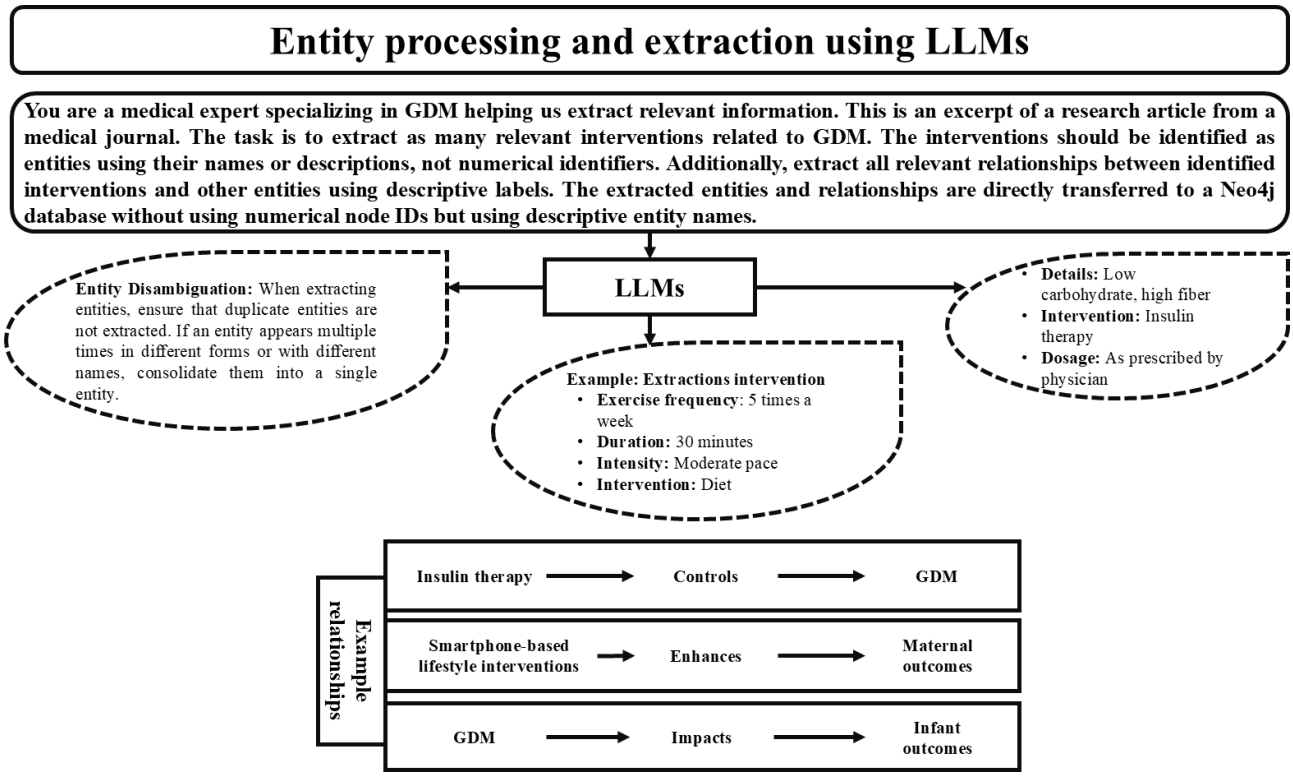
A few-shot prompting strategy was applied to guide the language model in identifying and structuring entities of interest in a usable format. Guided by 3 medical doctors, the prompts were manually engineered to show expected outputs, such as intervention types (eg, insulin therapy, diet, and physical 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].

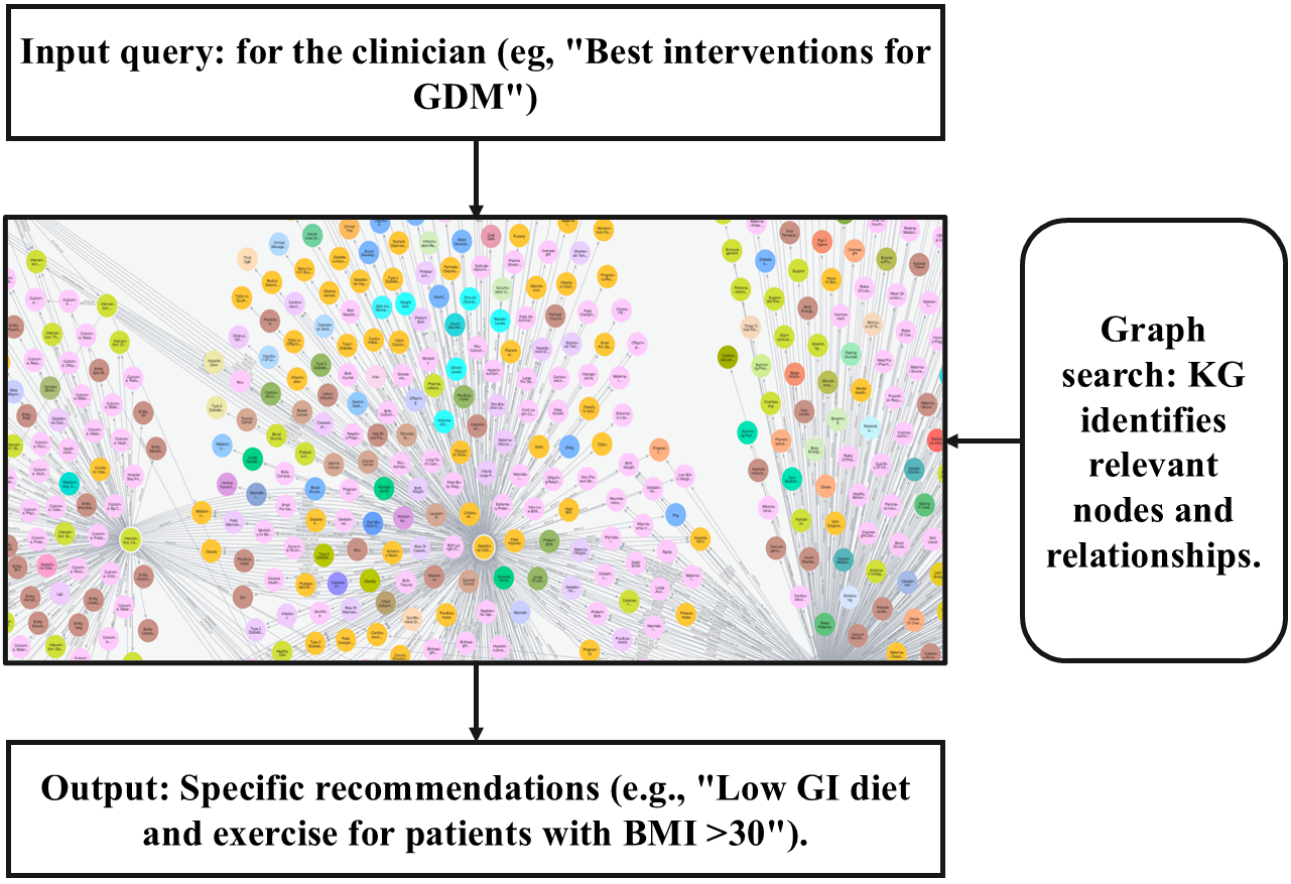
The KG construction process involved linking each extracted entity, such as insulin therapy, dietary strategies, or risk factors like obesity, to its semantically relevant mappings using directional edges labeled with relationship types (eg, "controls," "contributes to," and "enhances"). Each node was 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

(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.

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.

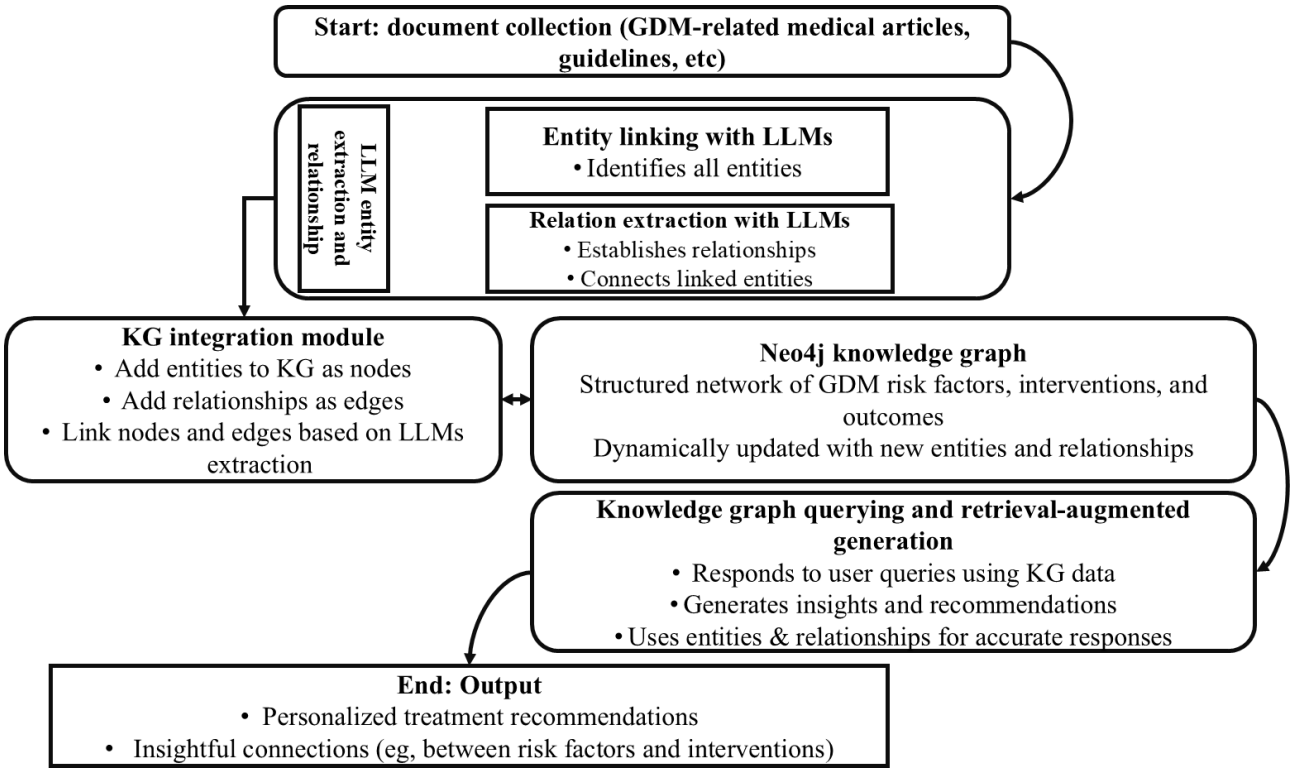


Figure 7. Structured retrieval and response generation process in graph-based retrieval-augmented generation. The diagram shows how clinician queries interact with medical knowledge sources, pattern matching, and graph-based retrieval to enhance artificial intelligence–generated responses. GDM: gestational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

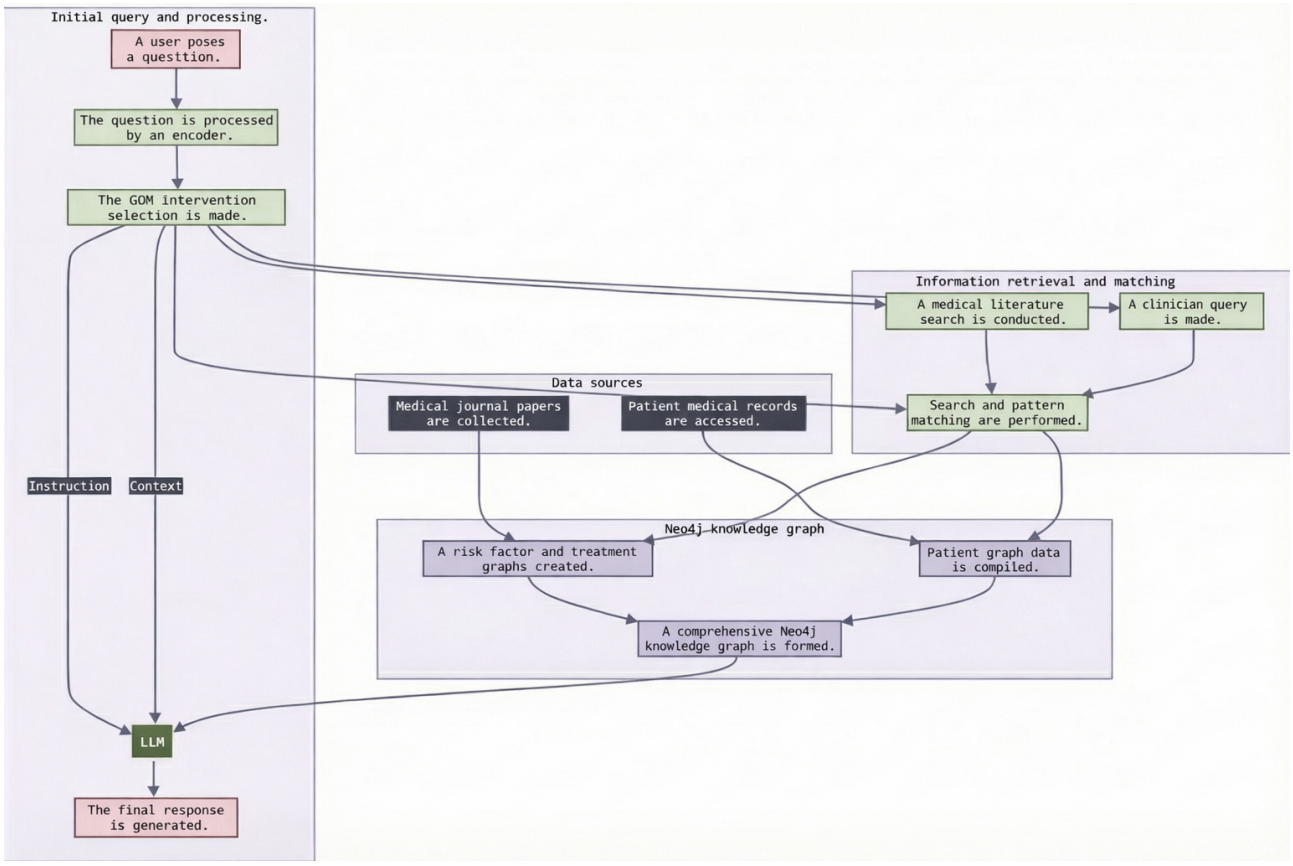


Figure 6 shows the underlying LLM-KG pipeline, including entity extraction, relationship linking, and graph query generation. Figure 7 offers a complementary perspective by emphasizing end-to-end data flow, from patient query and literature matching to LLM response generation, thus highlighting how both 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.

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.

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.

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

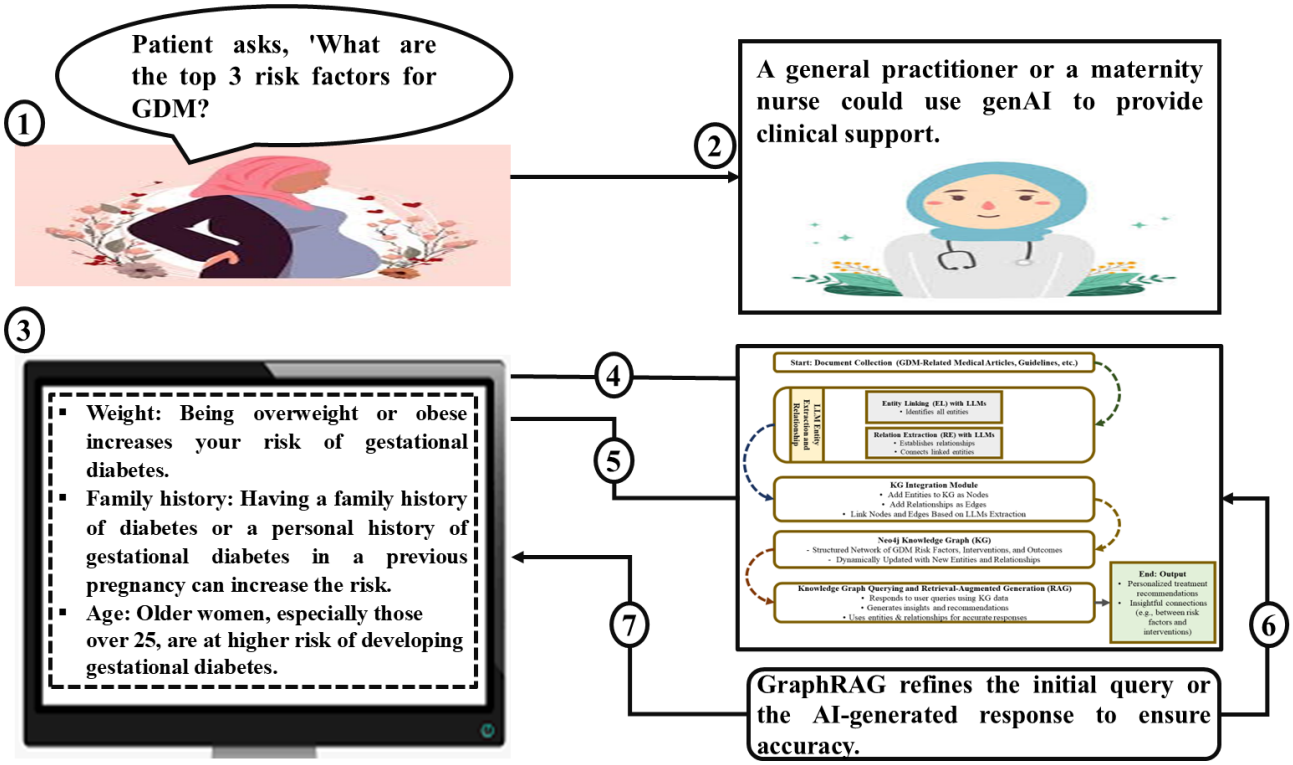
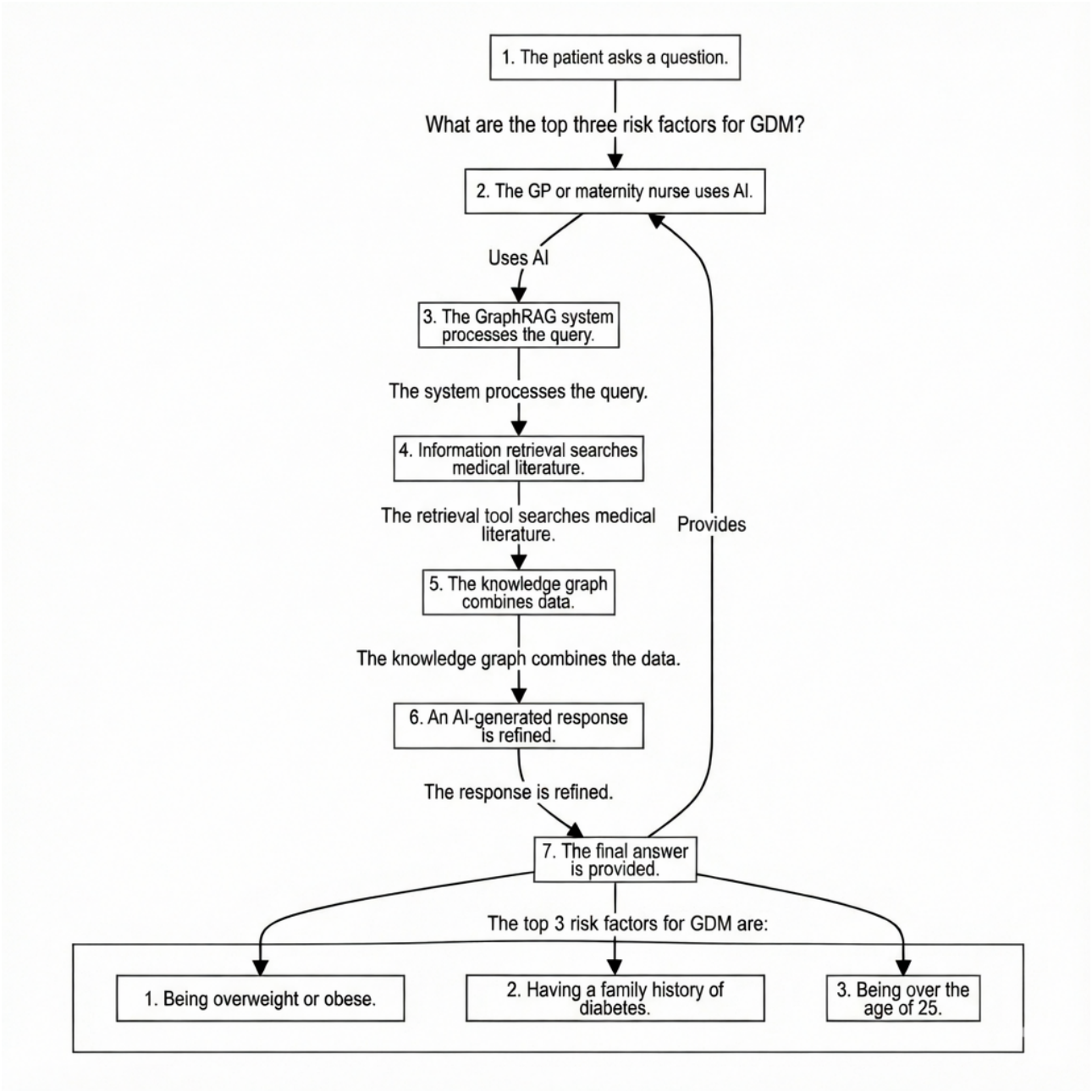


Figure 9. GraphRAG-based clinical support system for gestational diabetes mellitus - process flow diagram. AI: artificial intelligence; GDM: gestational diabetes mellitus; GP: general practitioner.



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.

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.

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.

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.

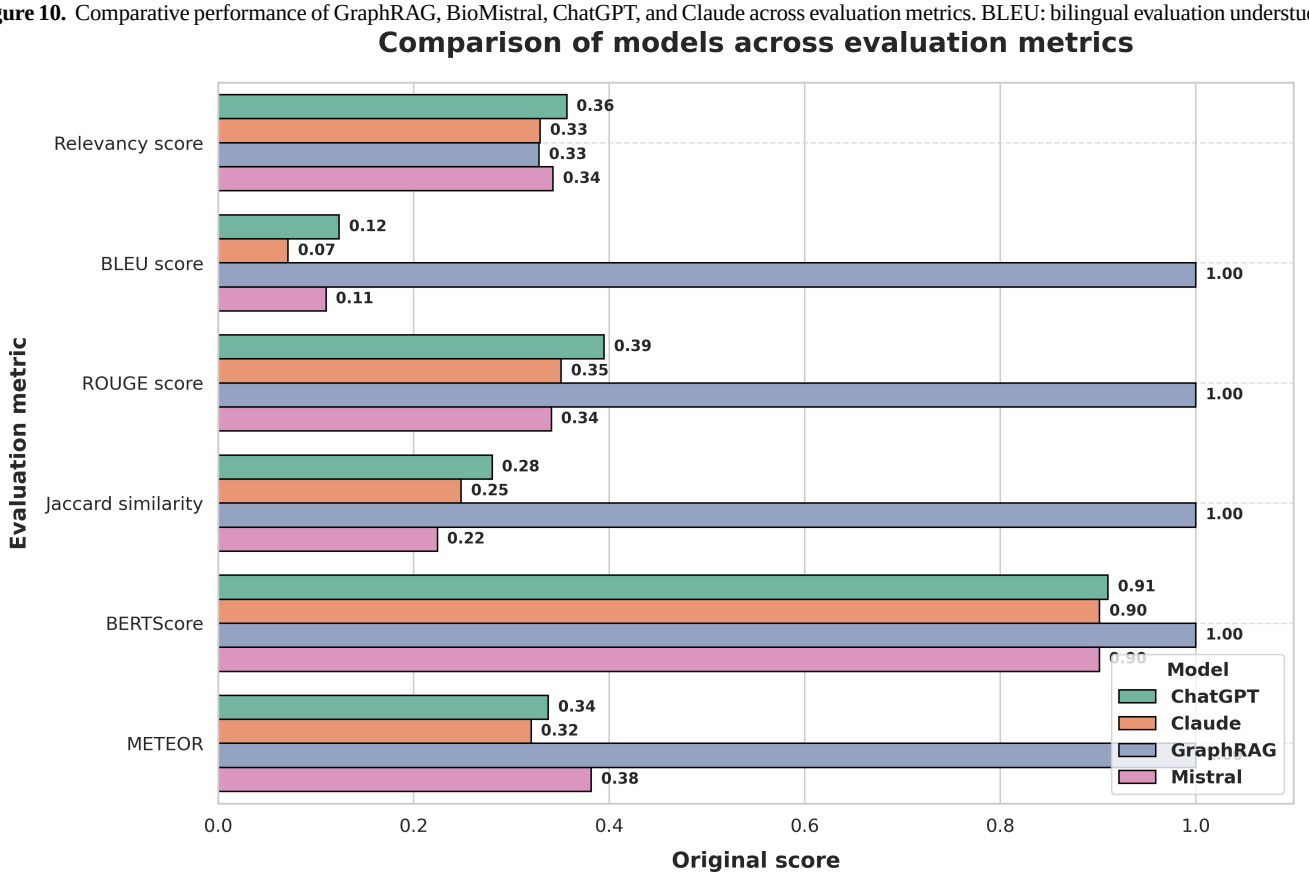


Figure 11. Radar chart visualizing model performance across key metrics. BLEU: bilingual evaluation understudy.

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

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,

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.

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.

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]

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]

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

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

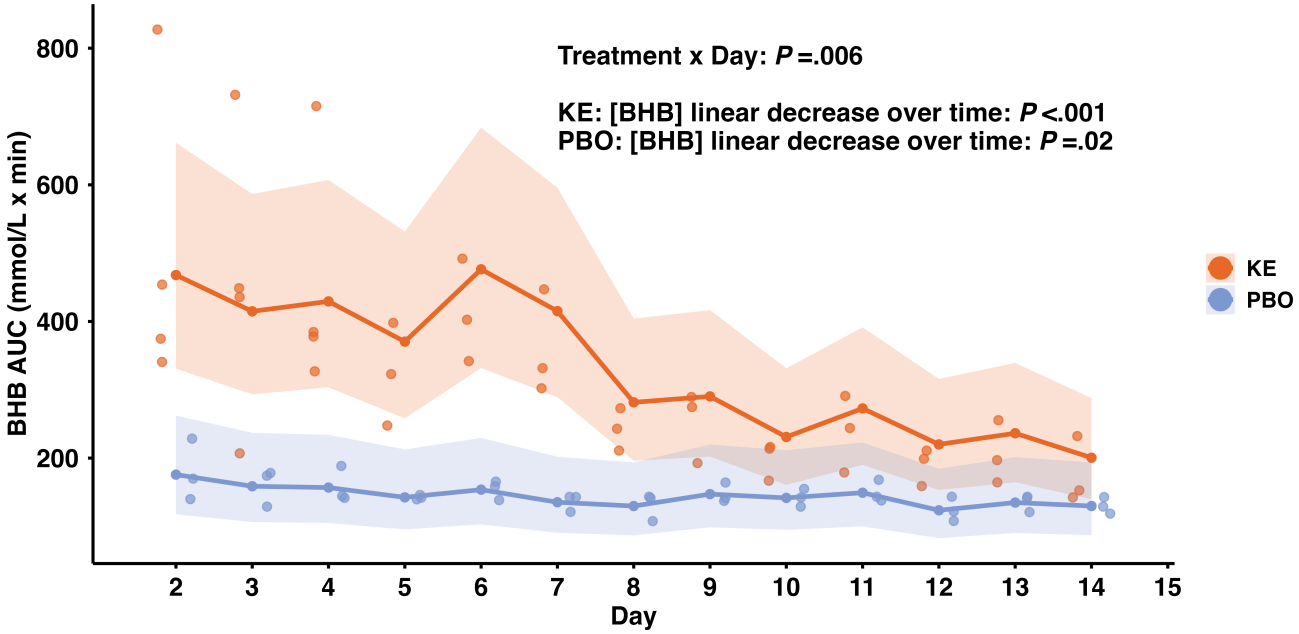
The study was conducted in accordance with the Declaration of Helsinki II, approved by the regional ethics committee (#1-10-72-221-22), and registered with ClinicalTrials (NCT06053138). Oral and written informed consent was obtained from all participating patients. Participant data were pseudonymized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

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.

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

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.

(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

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.

Figure 1. Flow diagram showing summary of the study procedures for the Weigh to Go intervention participants.

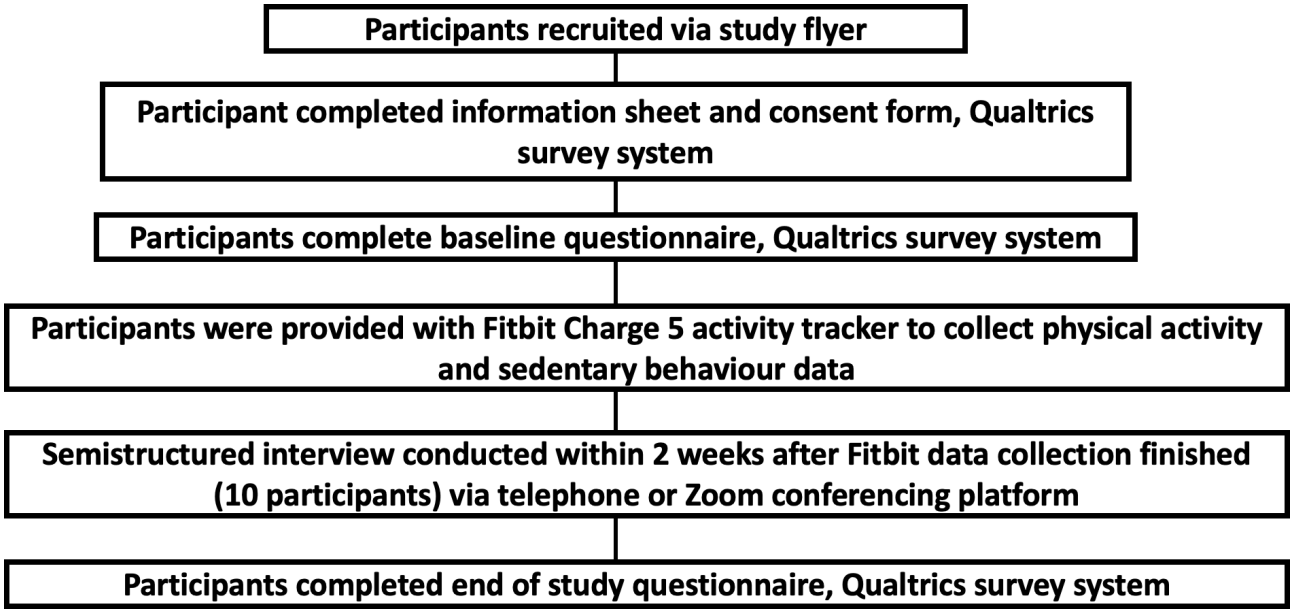
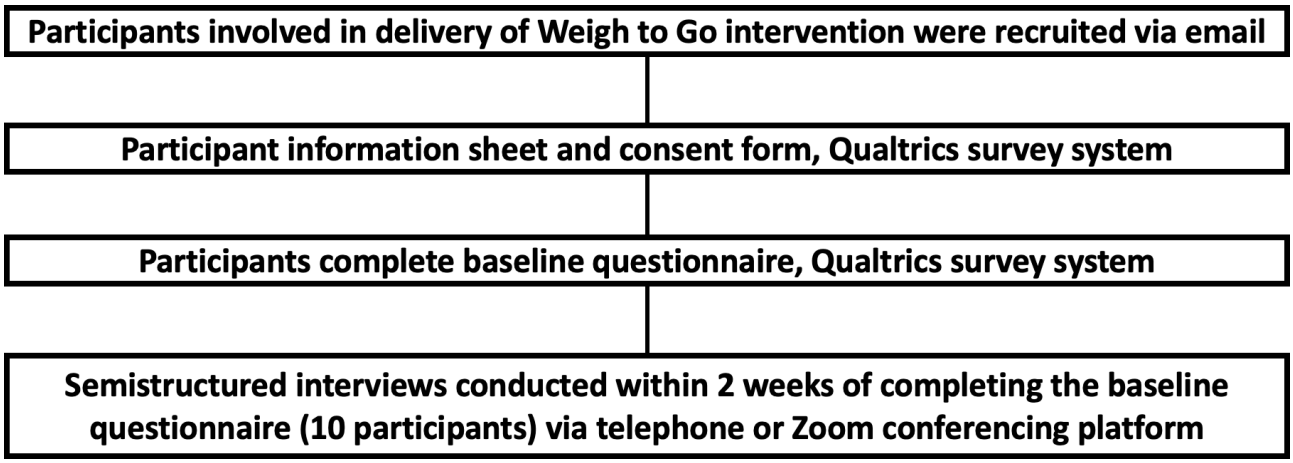


Figure 2. Flow diagram showing a summary of the study procedures for the Weigh to Go intervention for health care professionals.



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.

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

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

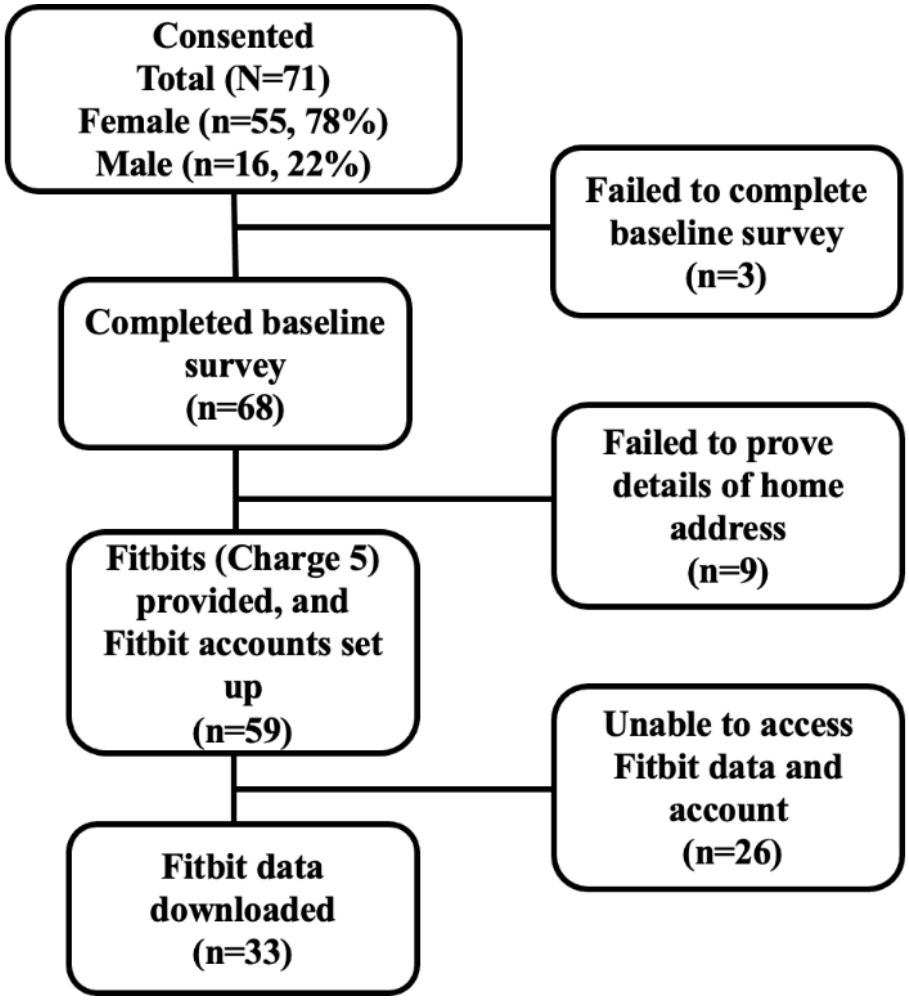
Ethical approval for this mixed methods study was obtained from the University of Strathclyde Ethics Committee (approval no 2454). For the study, both adults diagnosed with type 2 diabetes and health care professionals were provided with web-based participant information sheets. Informed consent was 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.

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.

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.

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)

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

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]

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]

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.

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.

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.

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

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]

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]

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

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

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

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

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).

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

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

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.

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.

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.

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,

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

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&search_scope=UoE&tab=Everything&docid=alma9923767330246&lang=en&context=L&isFrbr=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]

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

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].

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.

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

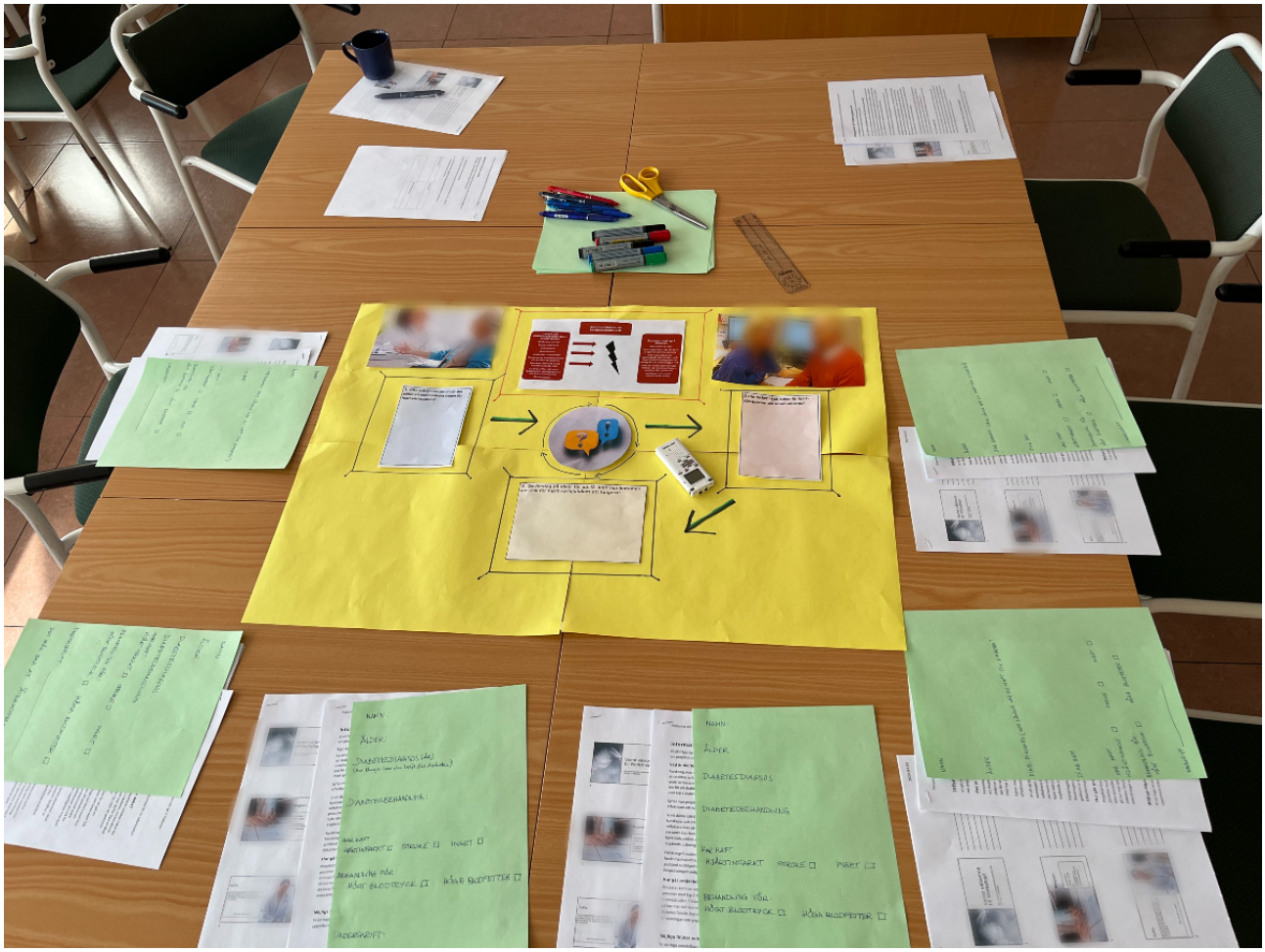
Data Collection

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.

Figure 1. Examples of probes and toolkits for the participants of workshops 1 and 2.



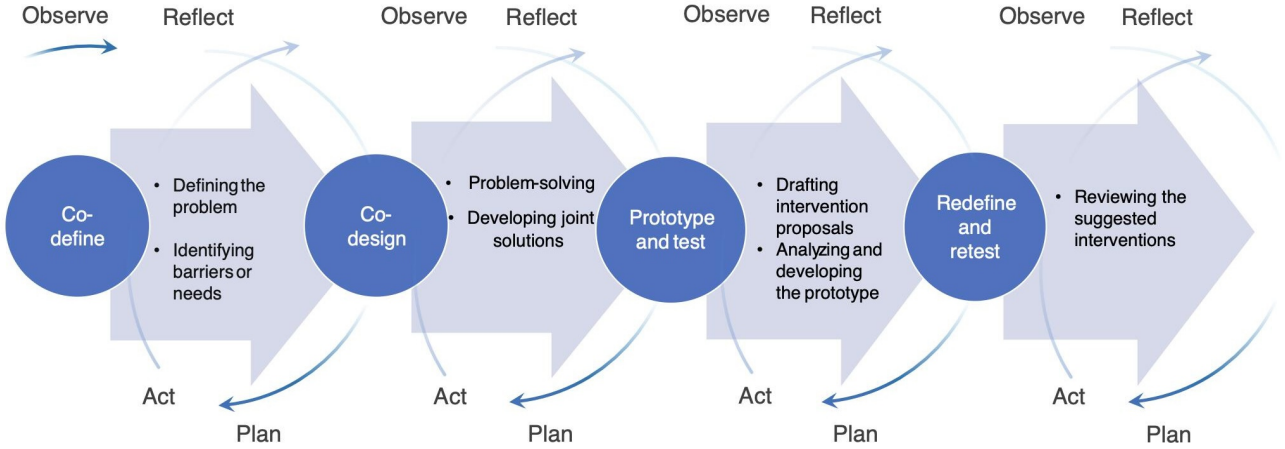
The Co-Creation Process Within a PAR Framework

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.

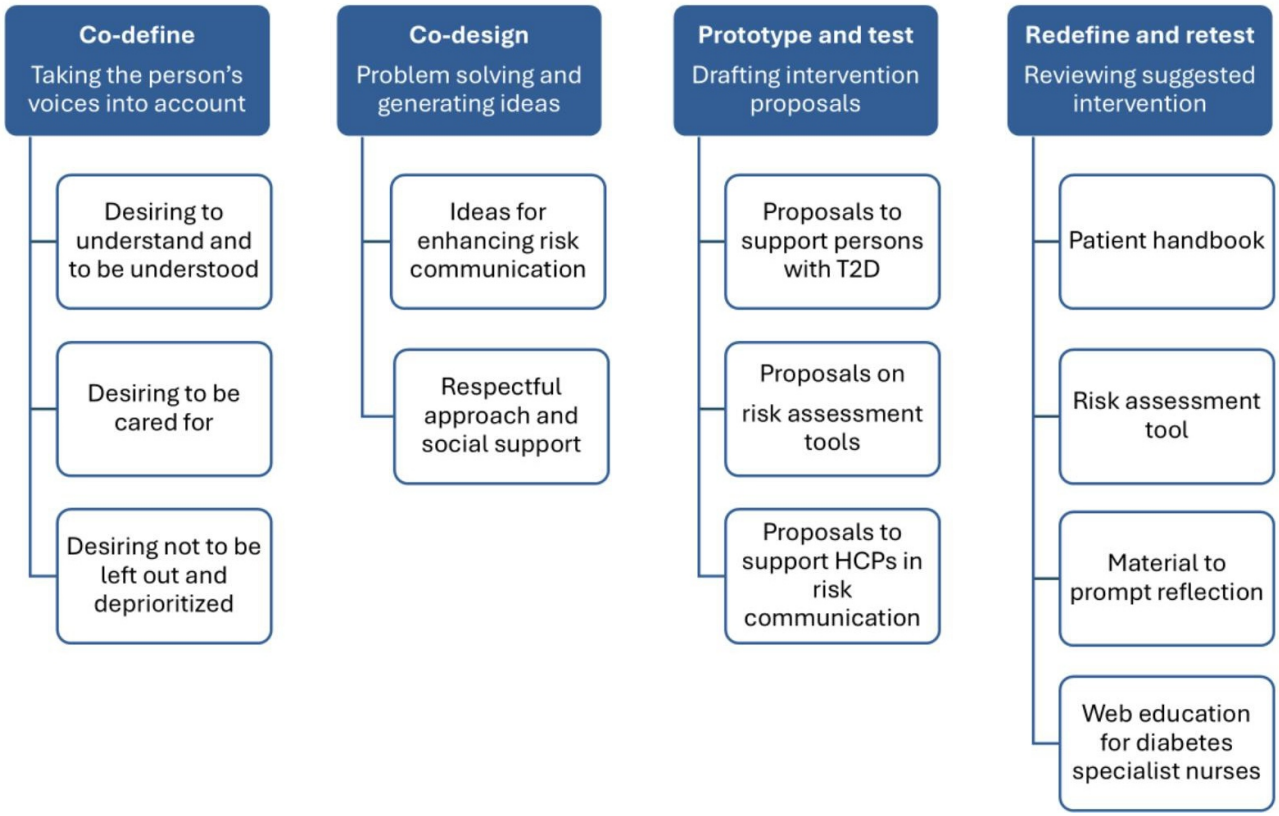
Figure 2. An overview of how the participatory action research (PAR) process proceeded and how the results evolved progressively over the course of the 4 co-creation cycles. The figure is influenced by previous research [16,19-21].



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).

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

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.

> *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

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.

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

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.

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

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]

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]

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

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

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

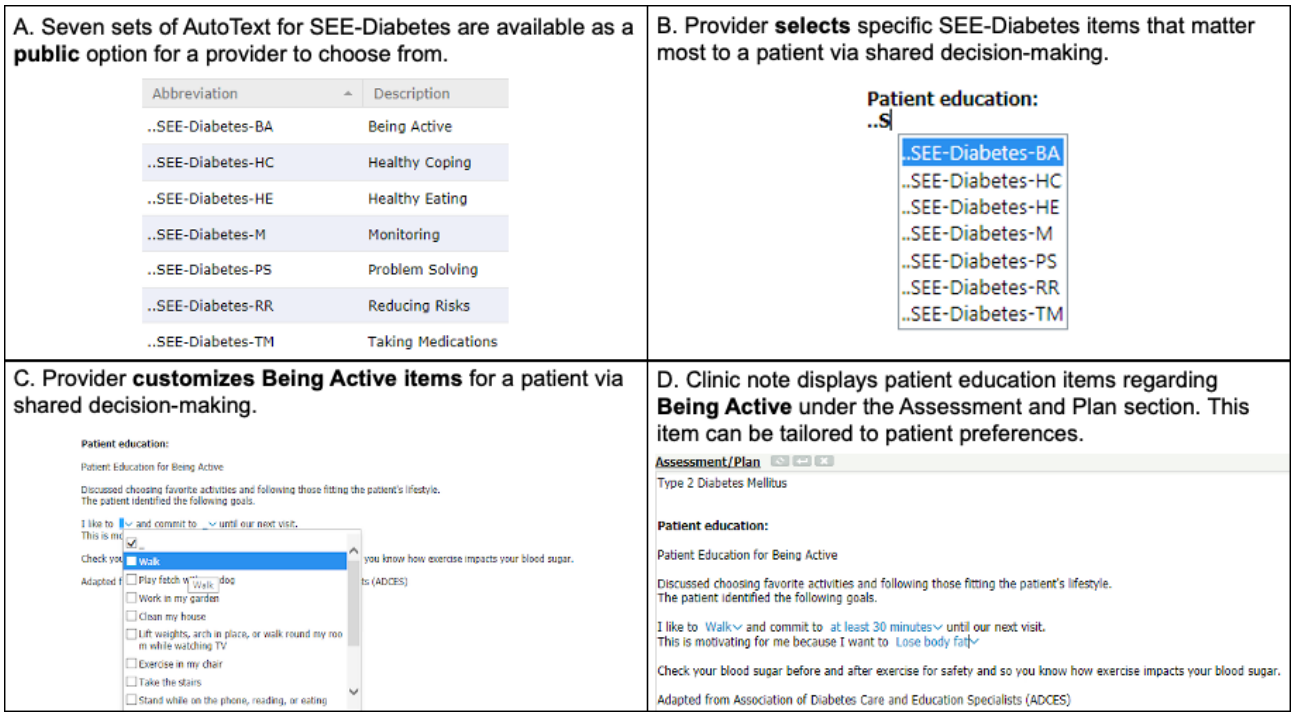
Study Design

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).

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.

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

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).

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

- 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).

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.

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).

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.

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

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

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]

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

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

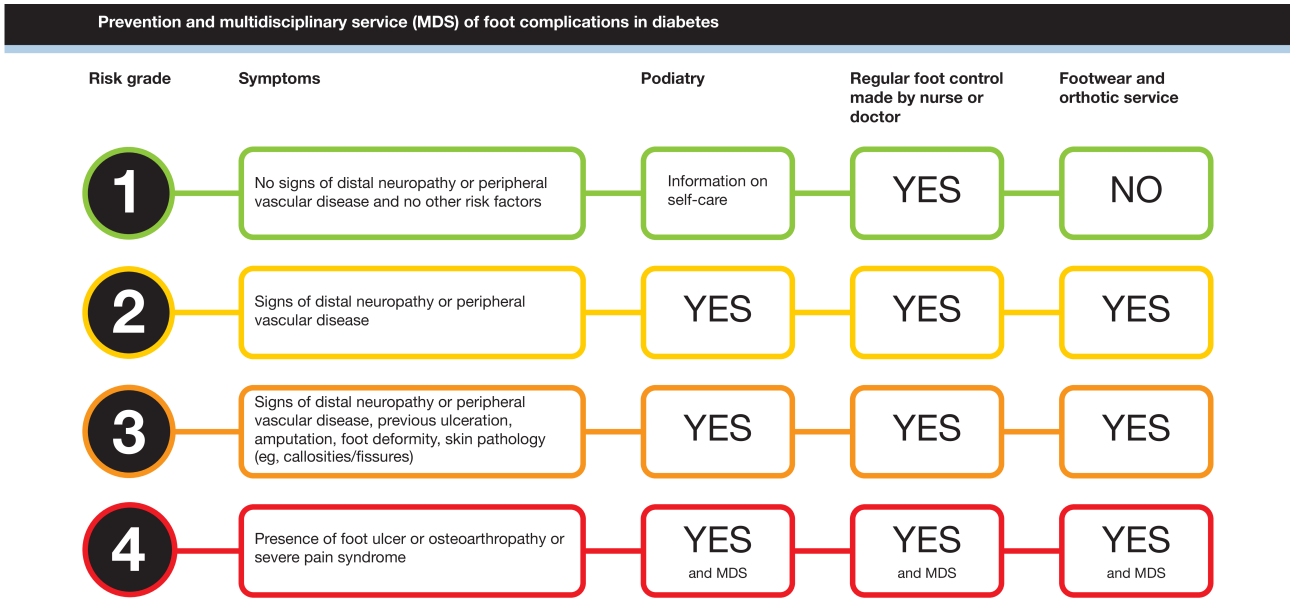
Introduction

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,

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,

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:

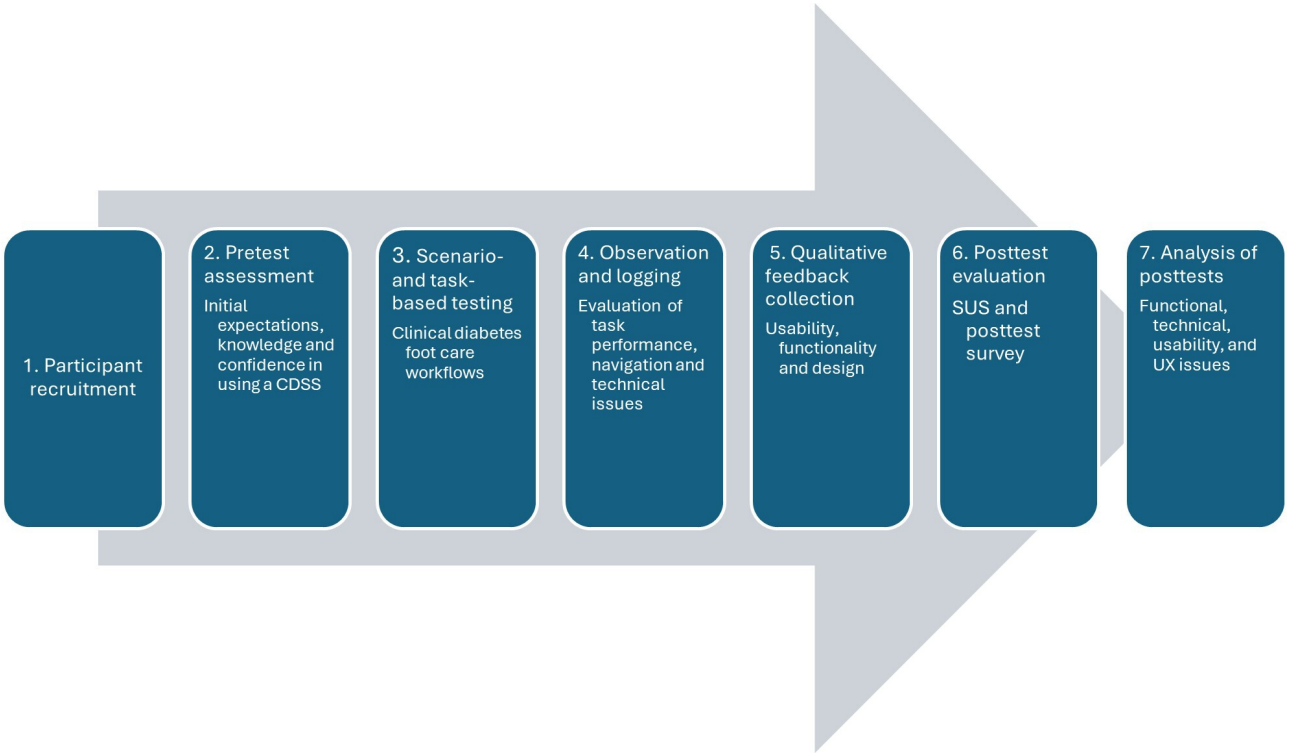
1. An innovation team, consisting of a clinical researcher, a researcher in health informatics, a master's student in biomedical engineering, and a researcher in biomedical engineering, formulated the strategy for the development of the digital health service.
2. The requirements for the digital health 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

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.

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.

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.

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

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.

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

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

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 Neubergh Foundation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

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]

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]

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]

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

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

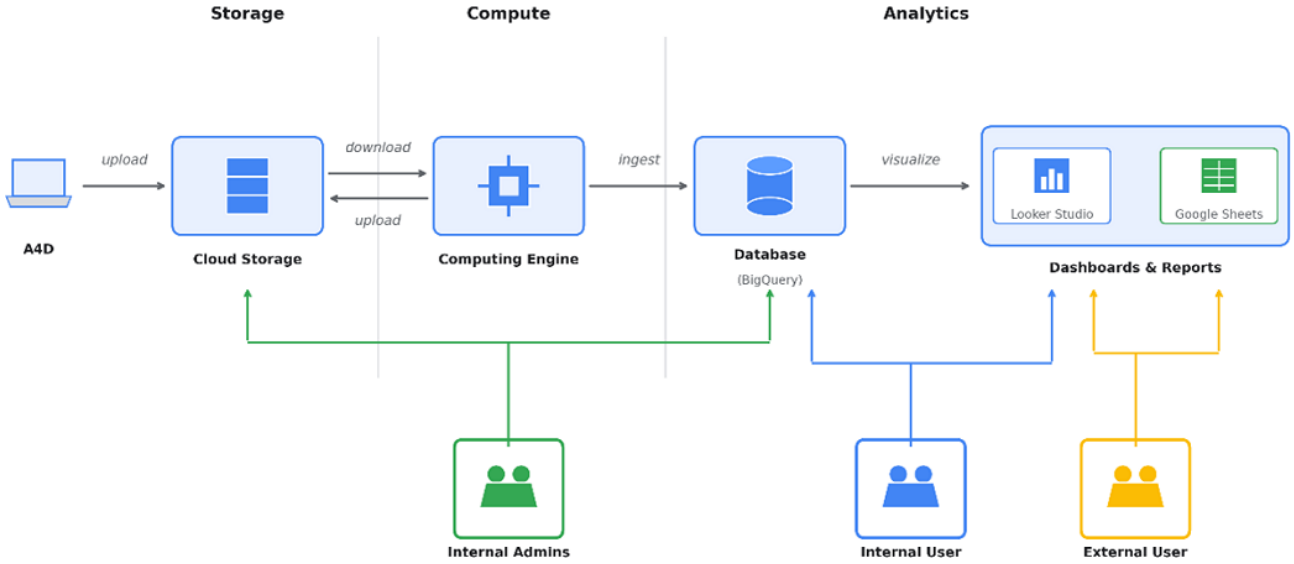
Recent decades have seen a dramatic proliferation of the use of real-world data and evidence generation from nonresearch settings, including both health and nonhealth sources [1]. Data utilization is particularly revolutionizing the operations and impact of not-for-profit, nongovernmental organizations 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

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.

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.

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

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

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

[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.

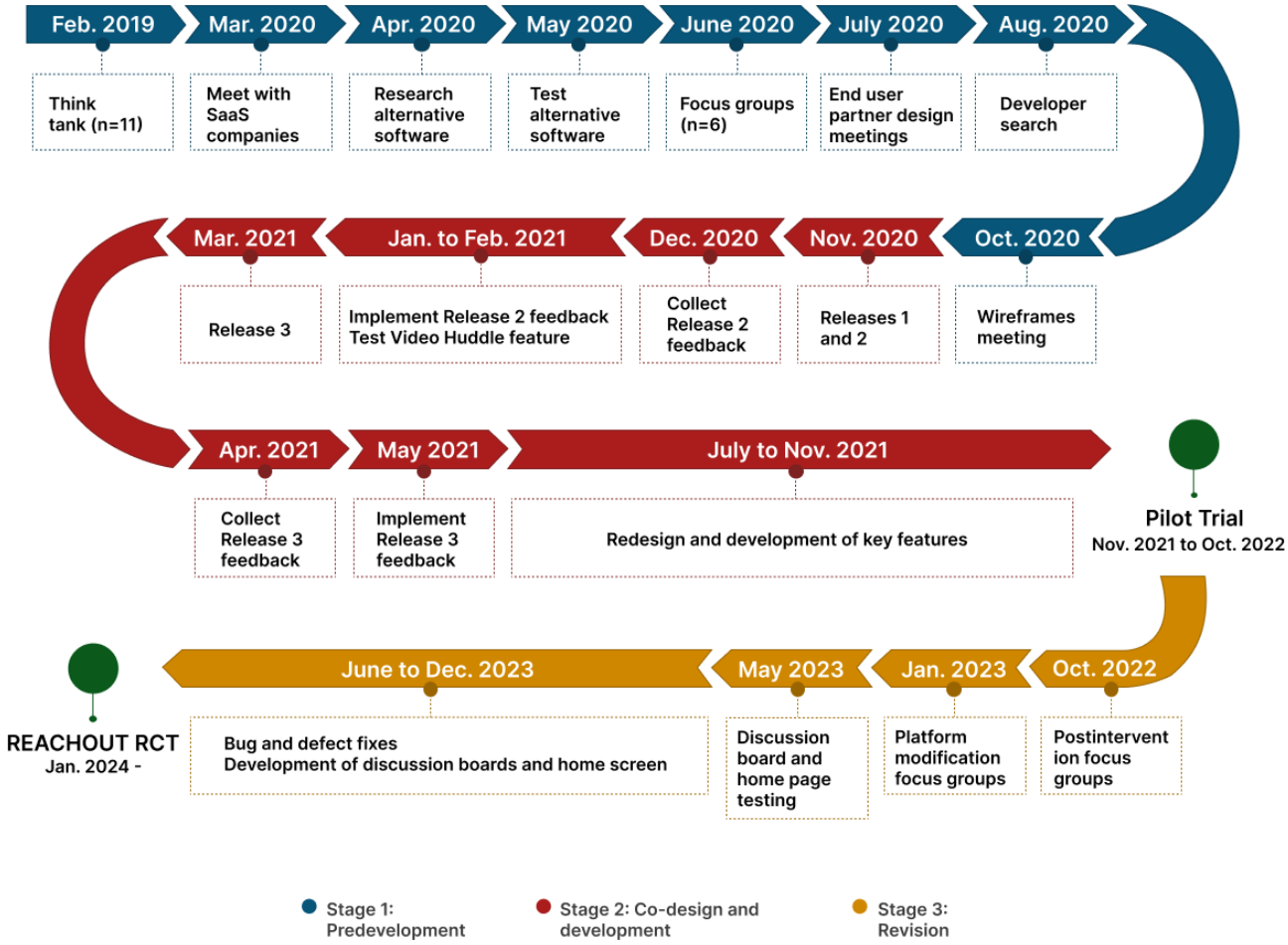
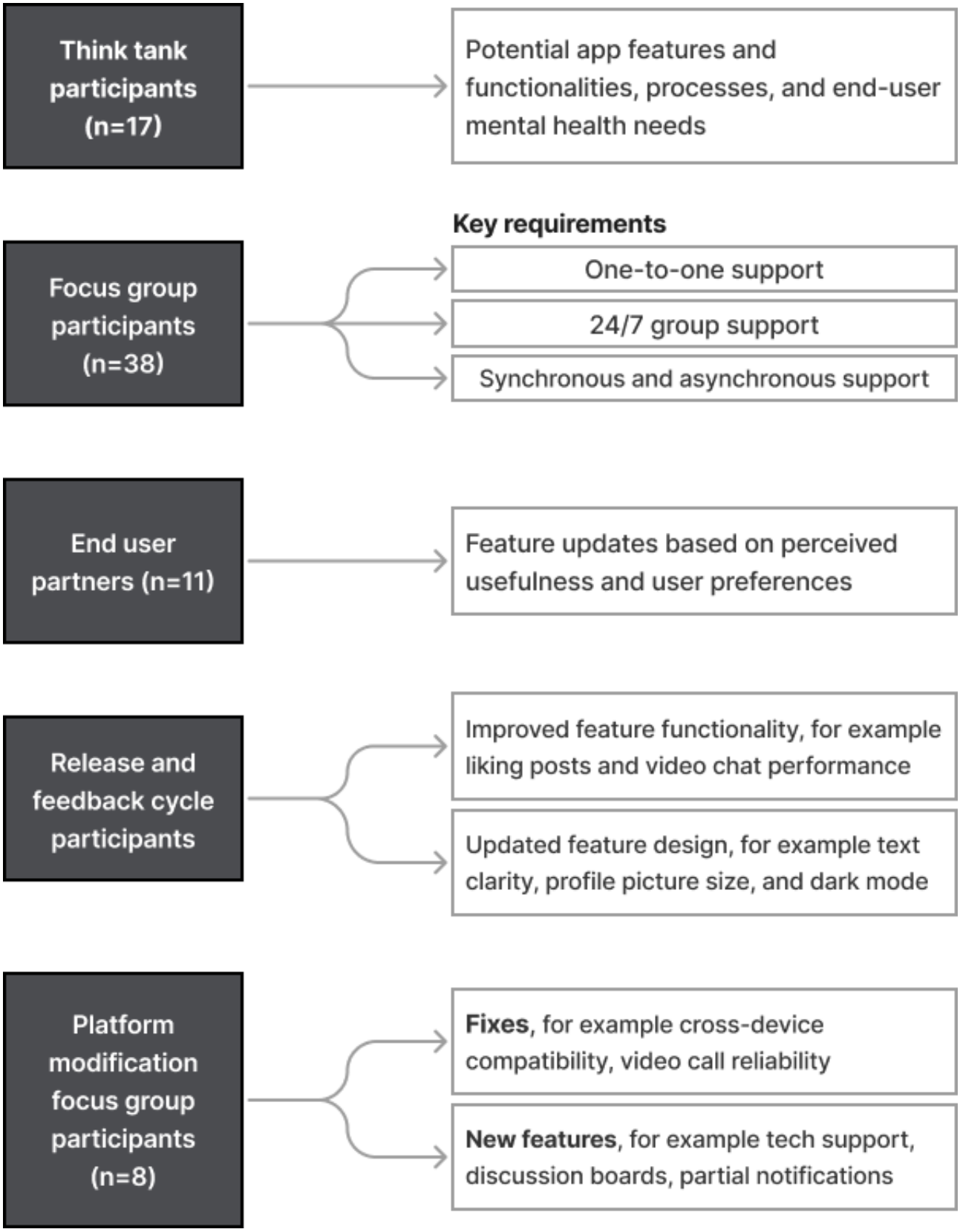


Figure 2. Co-design and development process, with each of the user groups engaged and the actions performed as a result of the activity.



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.

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].

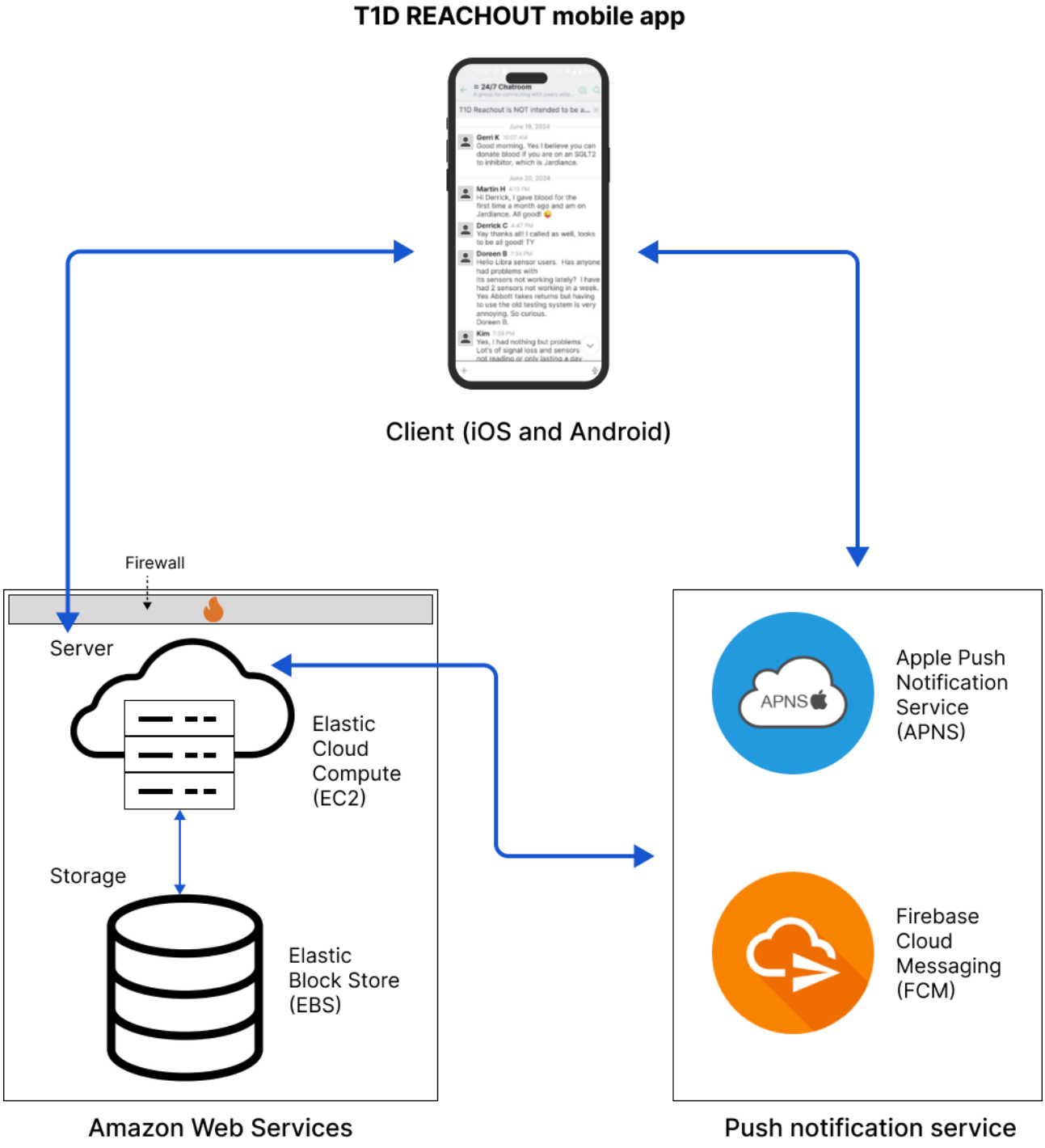
Property	Slack [27]	Rocket.Chat [28]	Mattermost [29]
Customizability	Slightly customizable	Very customizable	Very customizable
Pricing (per user)	CAD $207/year; US $ ~145/year	Free if self-hosted or CAD $55/year; US $ ~38/year	Free if self-hosted or CAD $141/year; US $ ~99/year
Hosting costs	Included	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).

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.

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.

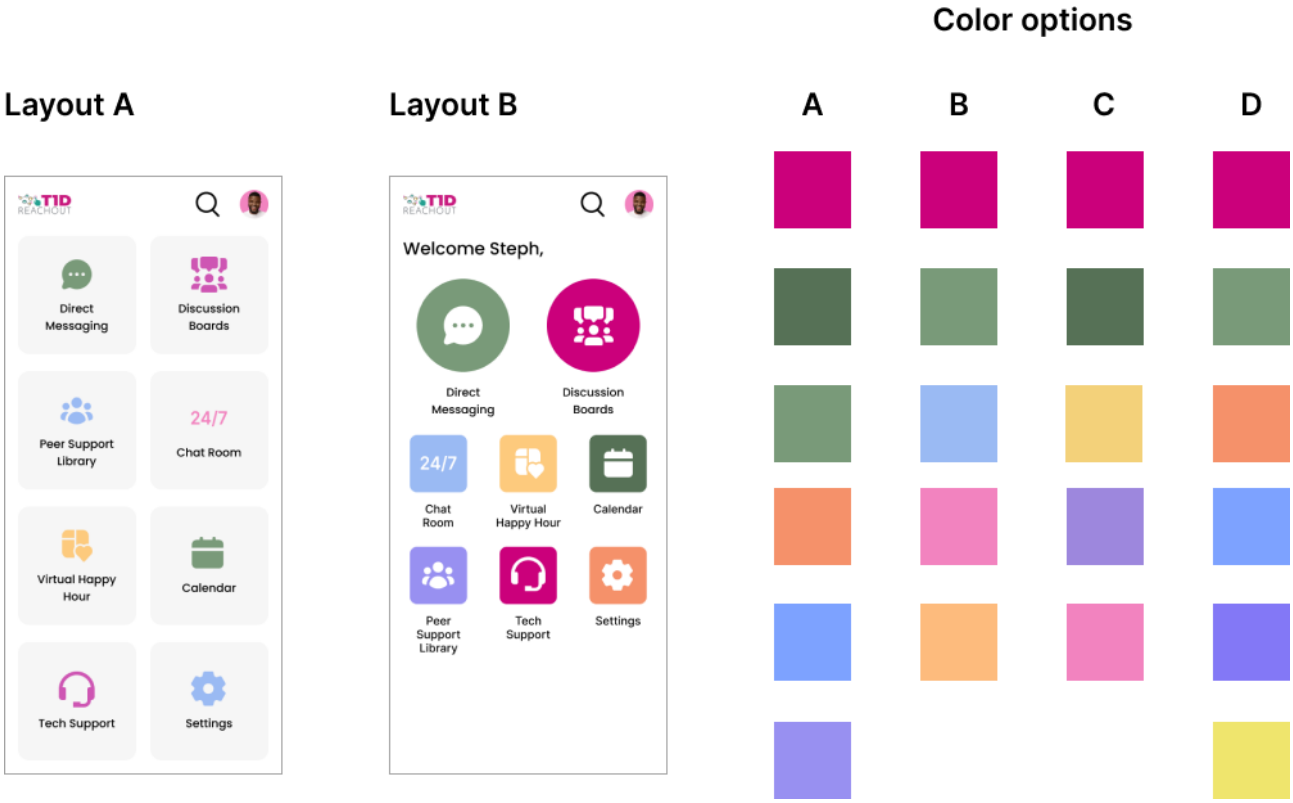
Platform Modification Focus Groups

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).

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

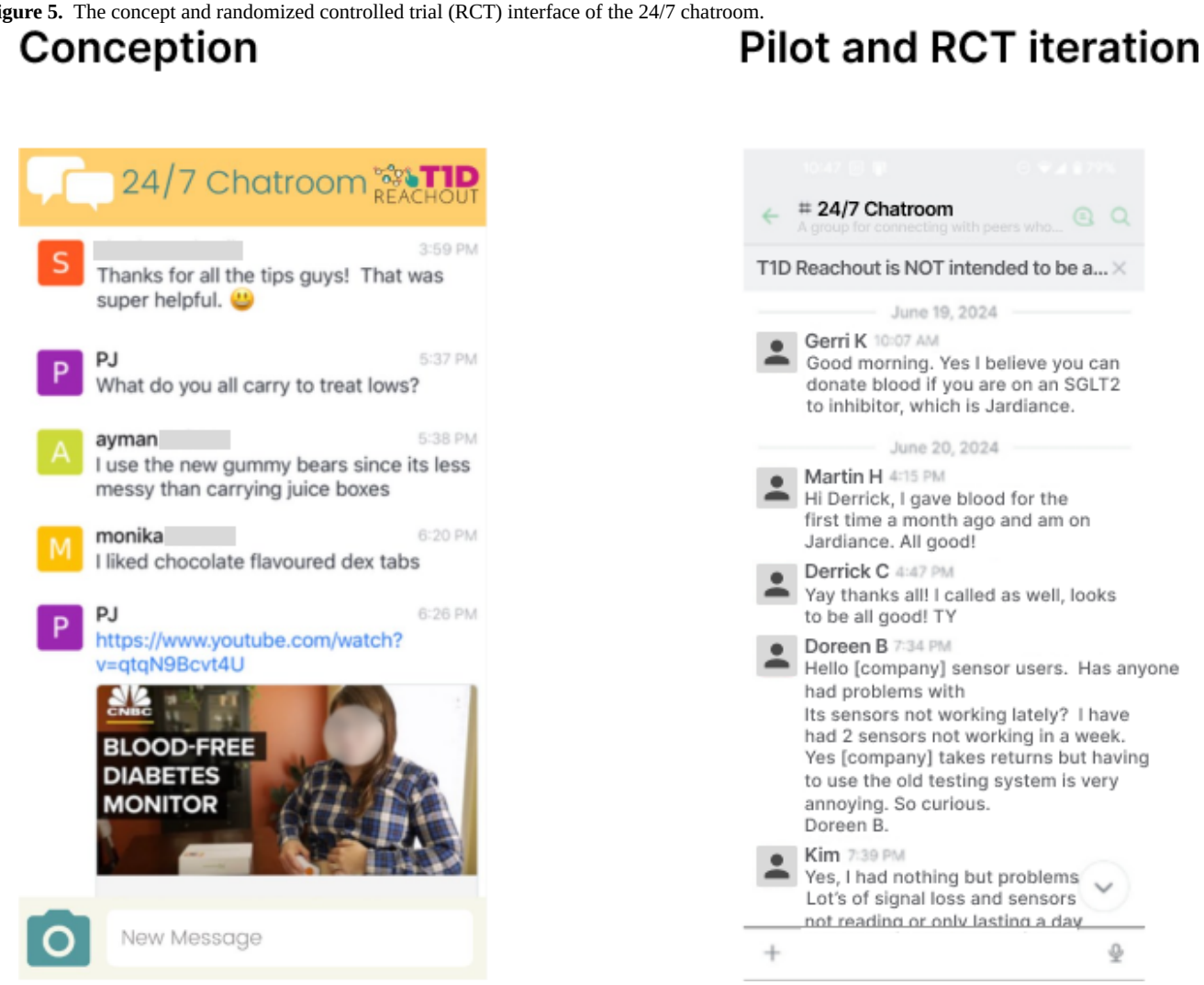
Overview

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).

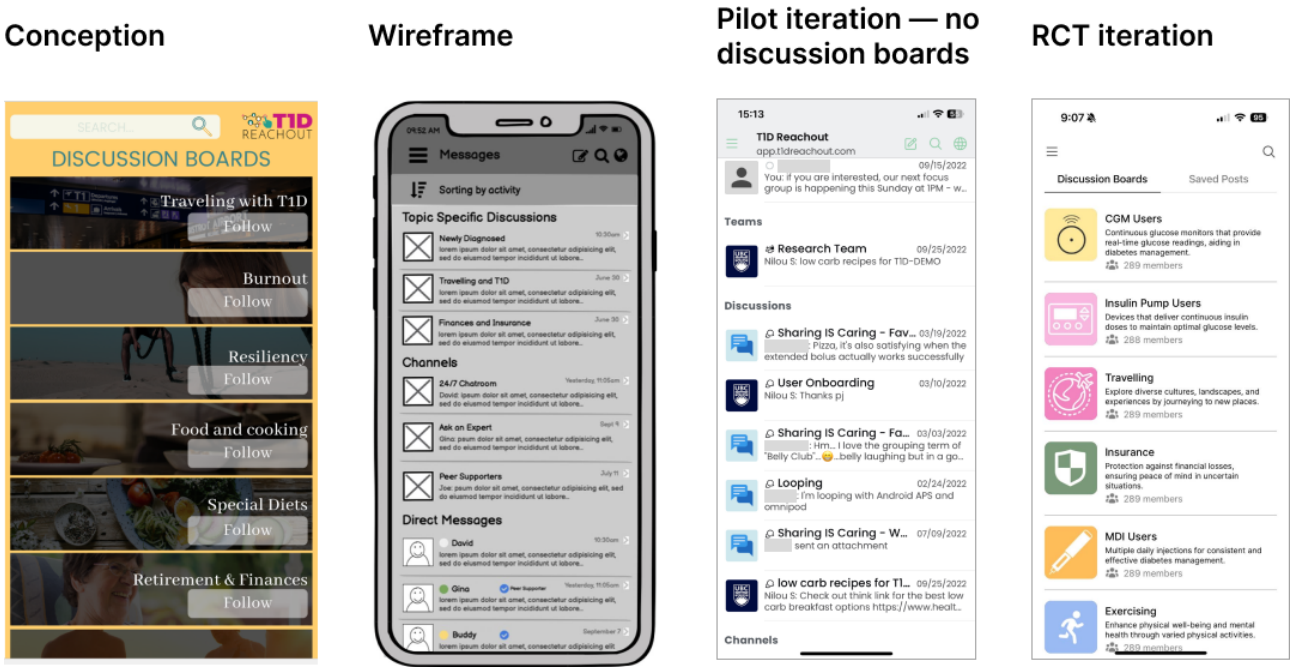
Figure 5. The concept and randomized controlled trial (RCT) interface of the 24/7 chatroom.



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.

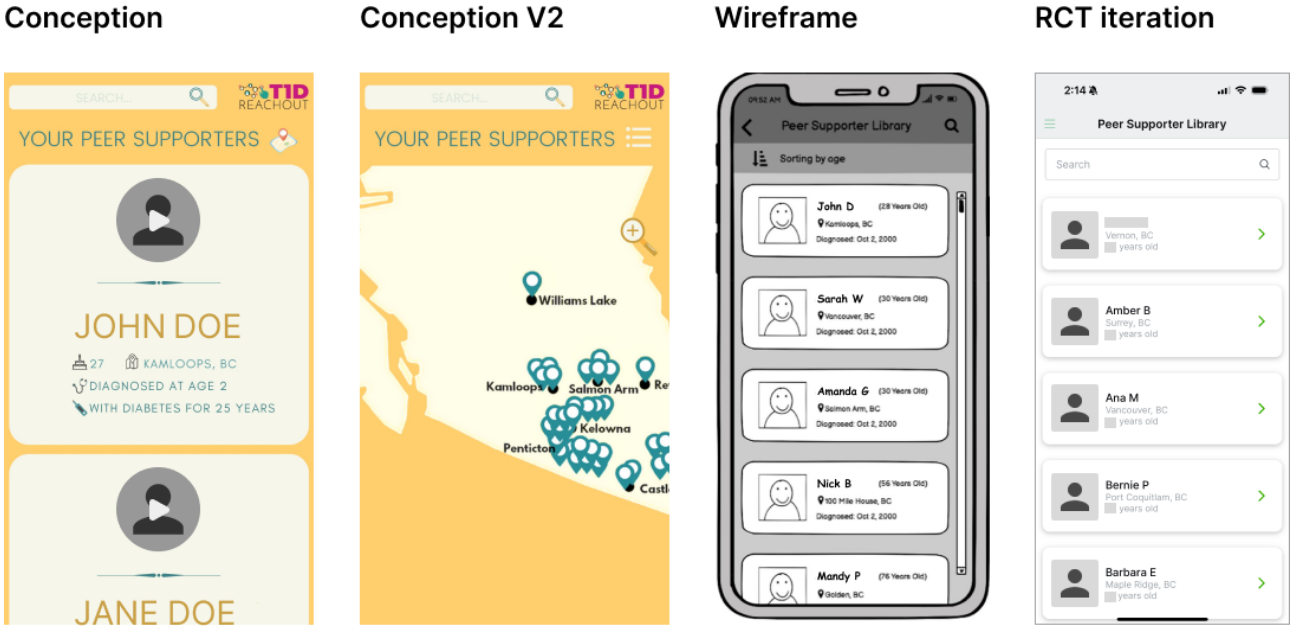
Figure 6. Discussion board iterations, including conception, wireframe, pilot iteration, and the randomized controlled trial (RCT) iteration, based on user feedback from the co-design and development process.



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.

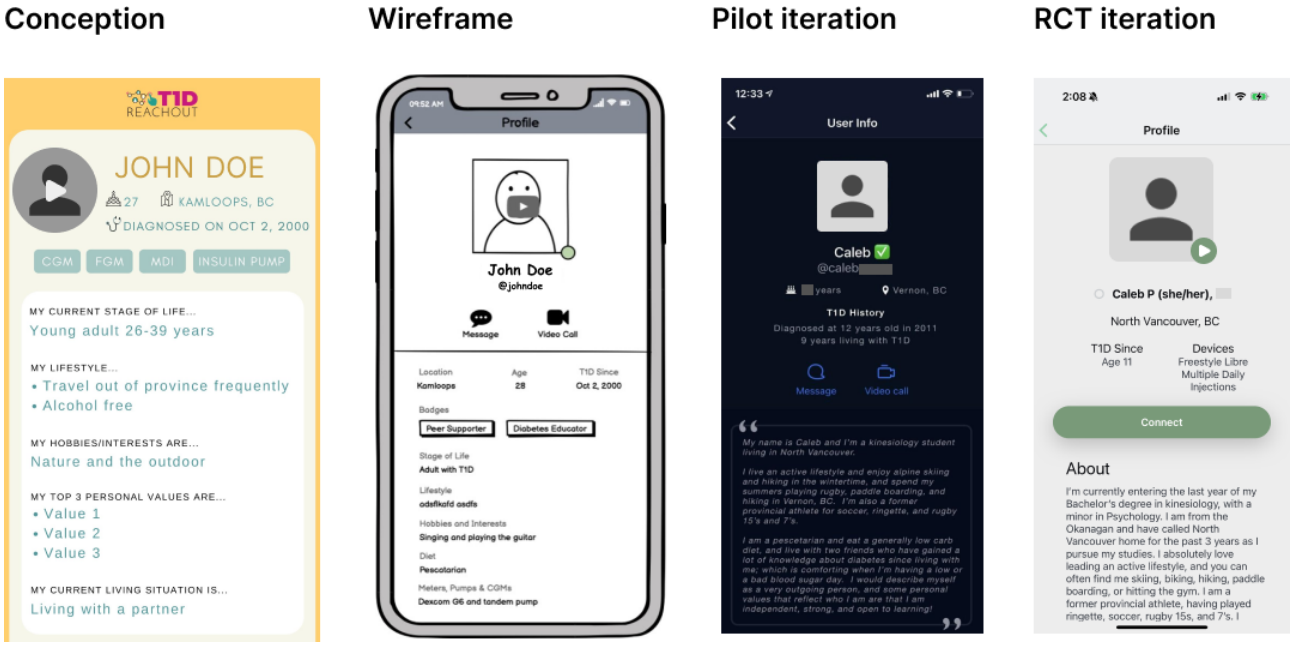


Peer Supporter Profile

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

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.



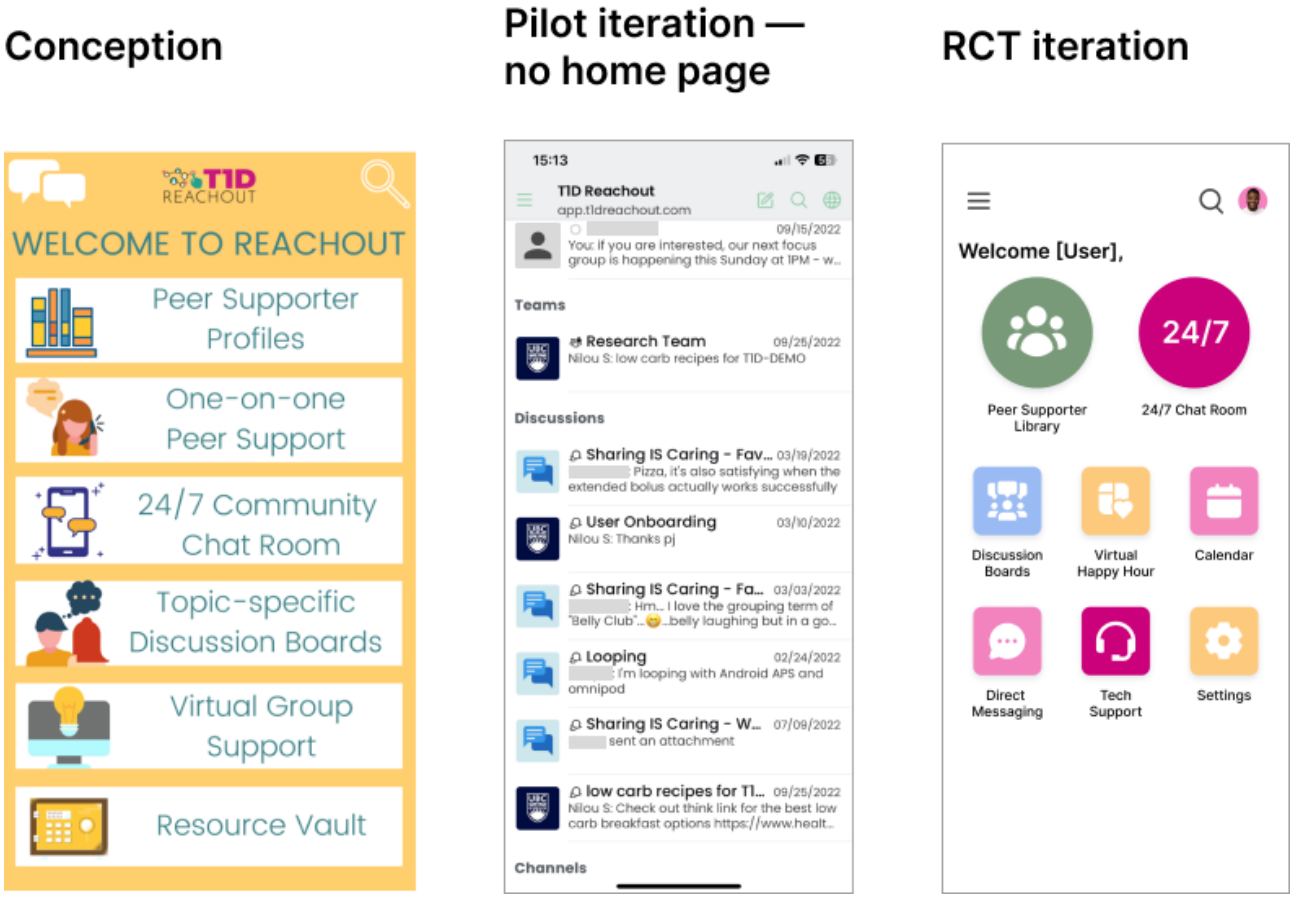
Small Group Virtual "Happy Hour" Sessions

"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.

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.



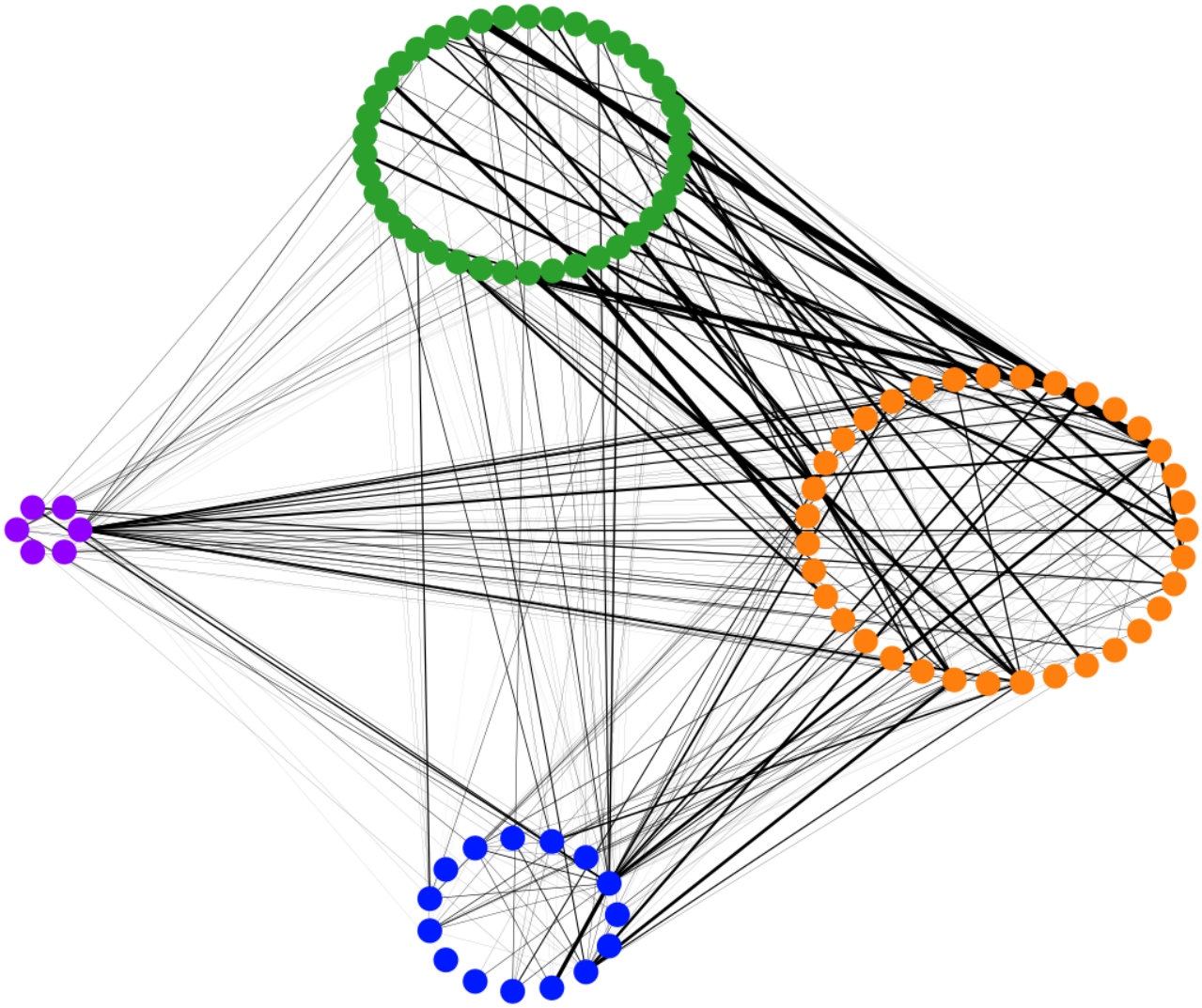
Initial Platform Utilization Patterns

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.

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.

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

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]

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

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

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

(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

(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

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.

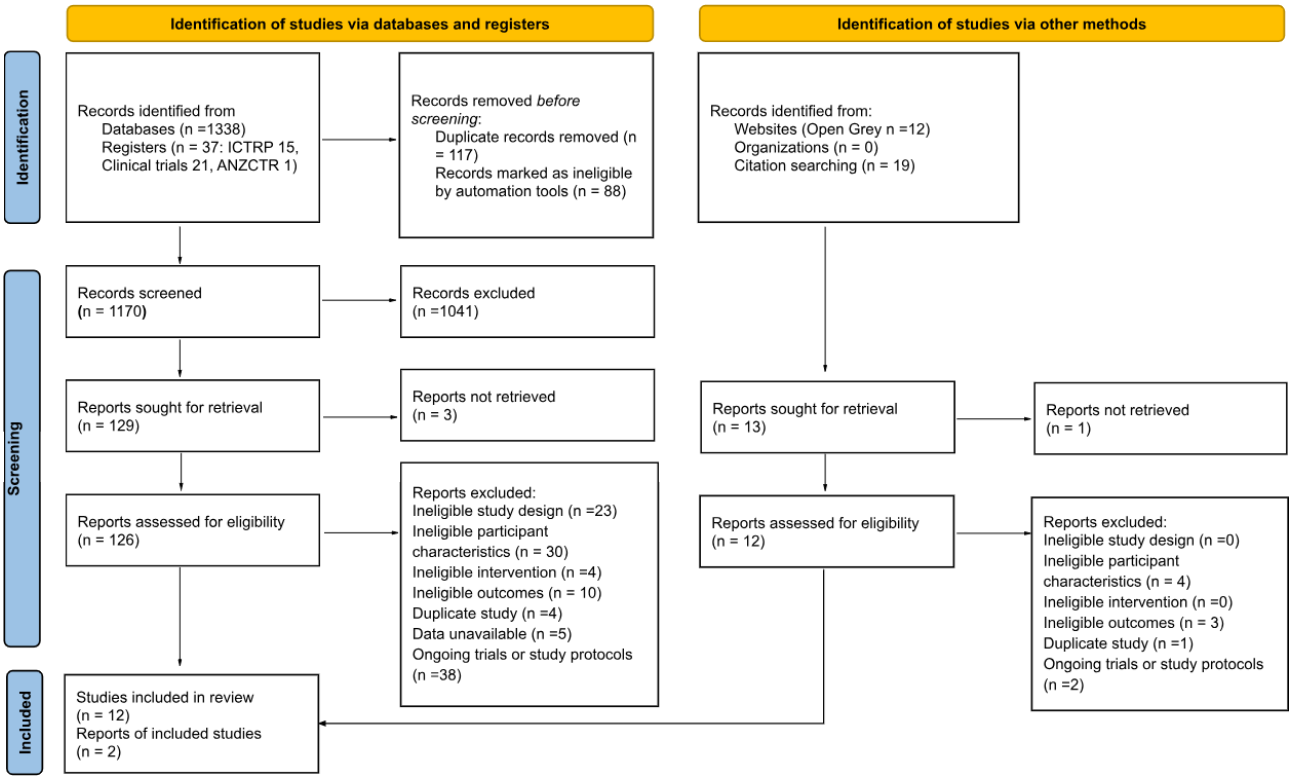
Results

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.

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.

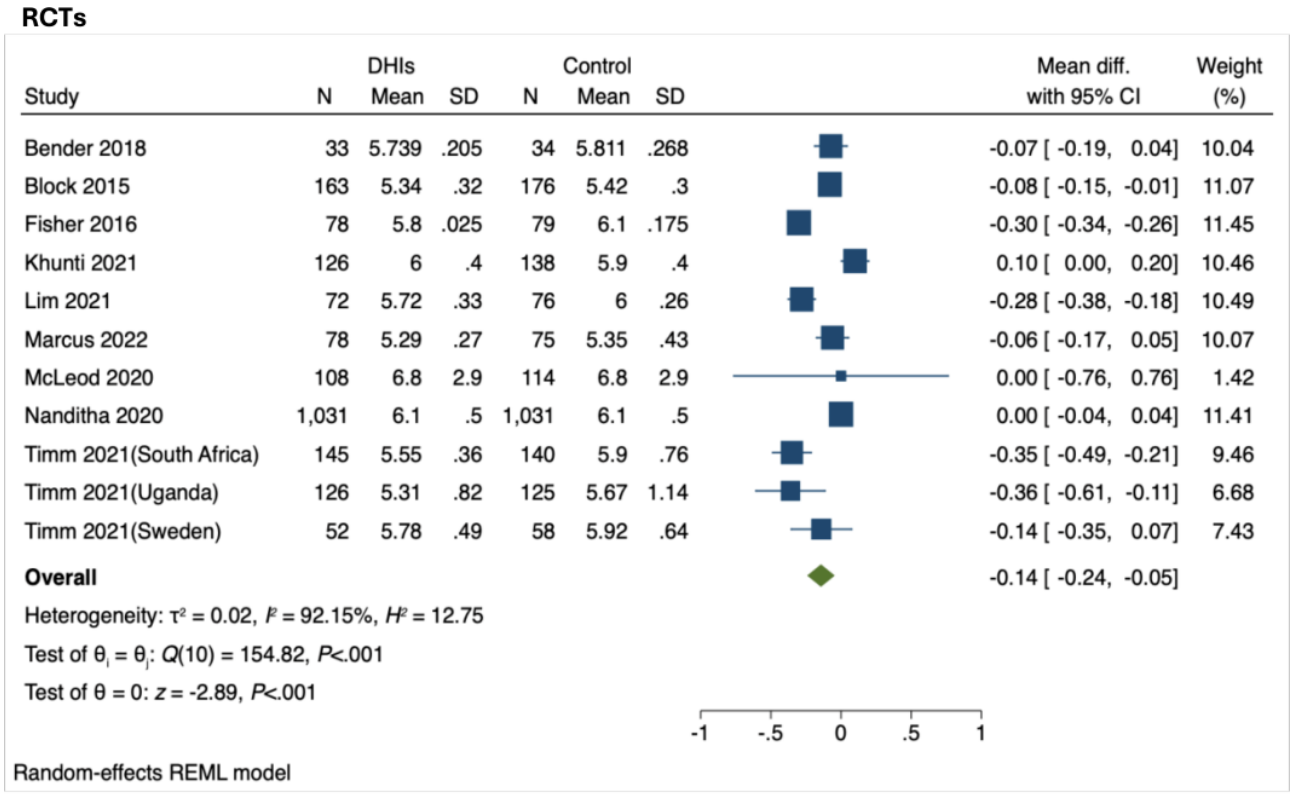
Primary Effectiveness Outcomes

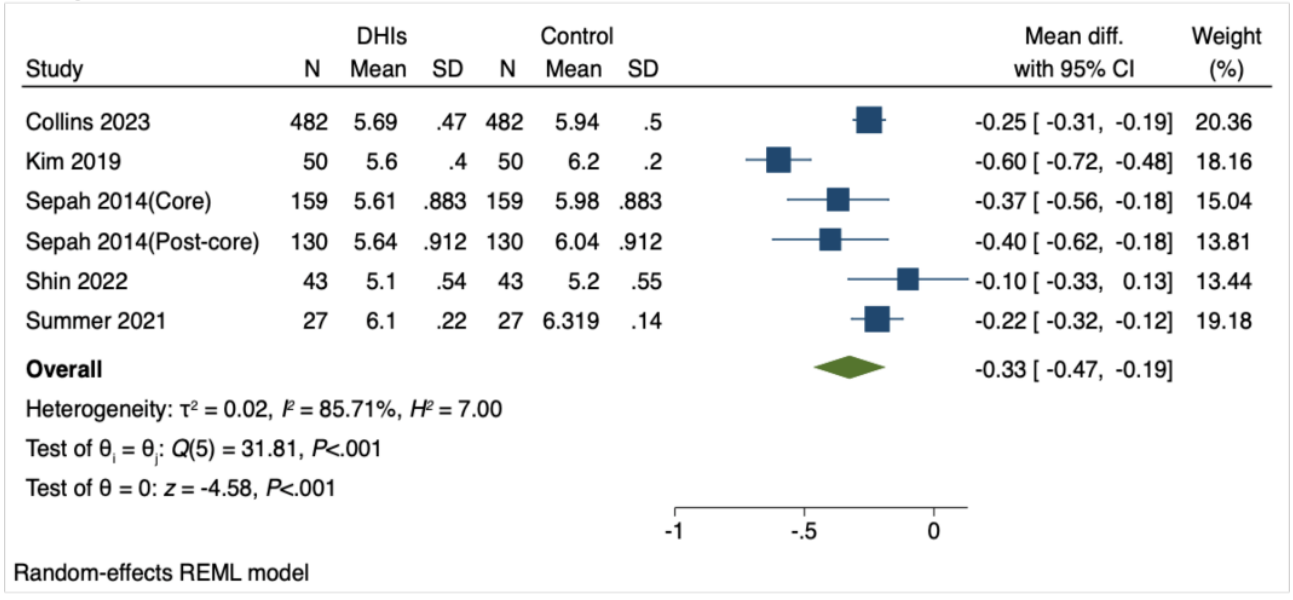
As Cochrane recommends that randomized trials and nonrandomized studies of interventions should not be combined in a meta-analysis [29], the meta-analyses for RCTs and pre-post studies were conducted separately.

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.

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

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.

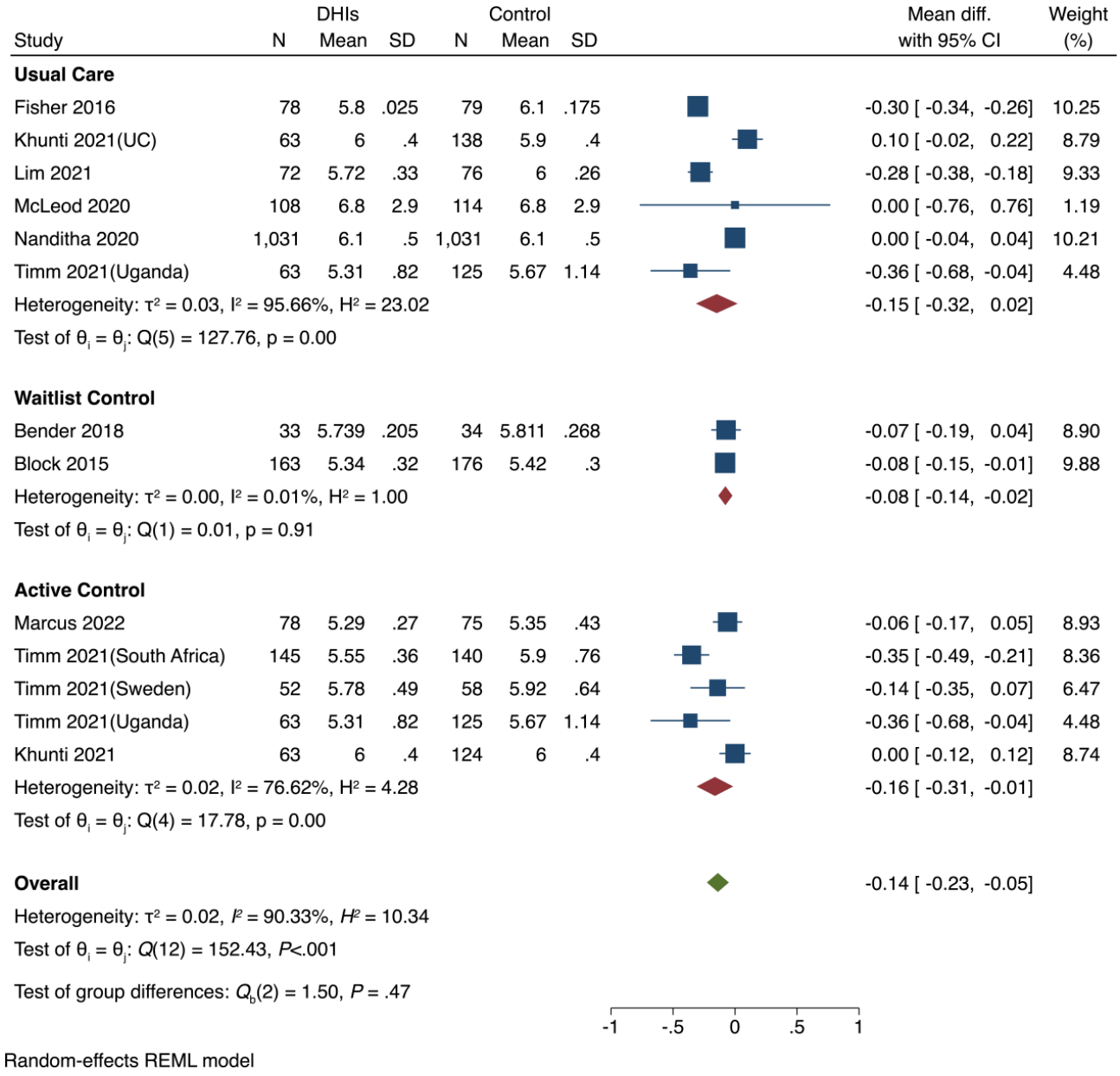
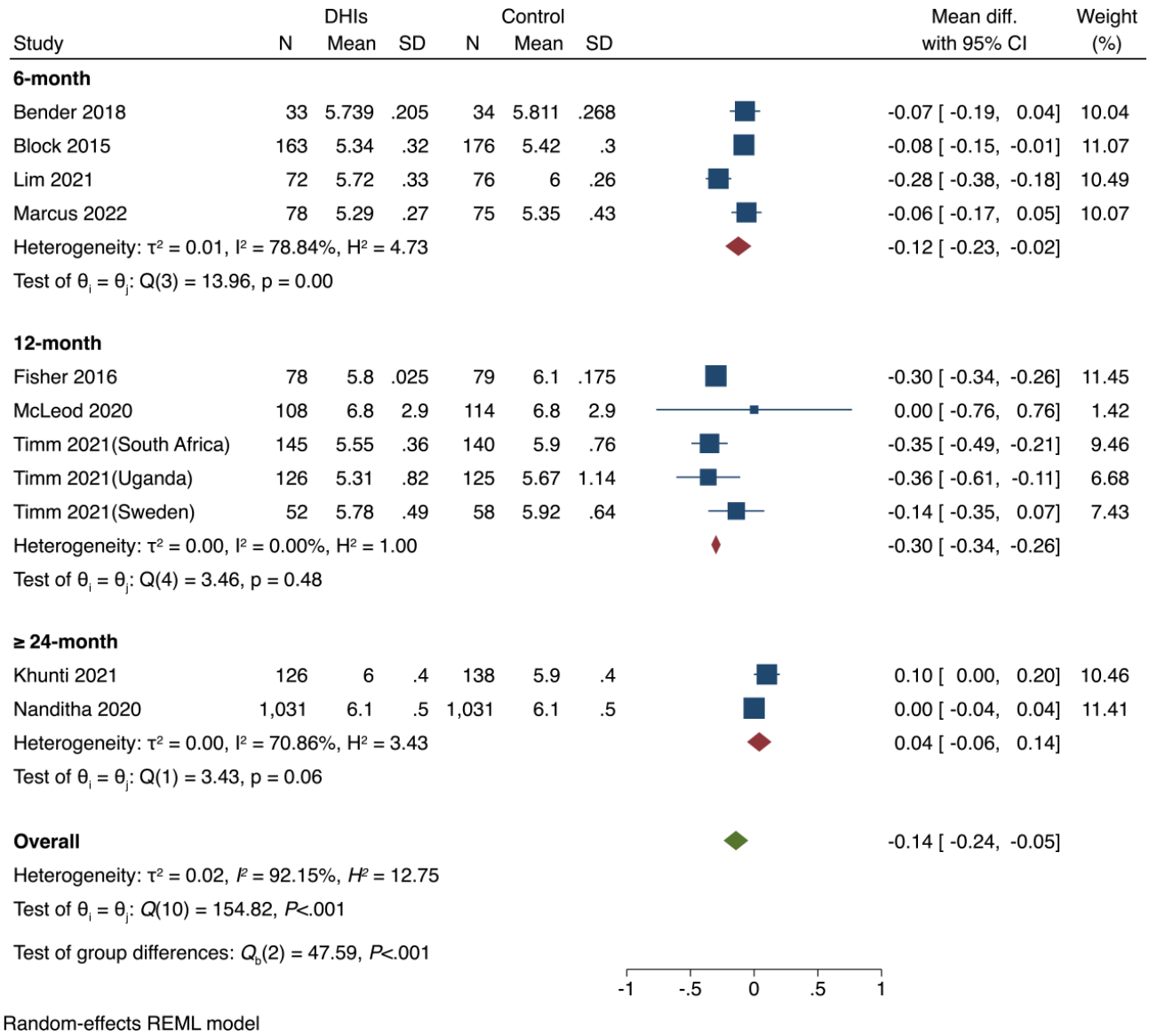


Figure 4. Subgroup analysis of hemoglobin A_{1c} (HbA_{1c}, %) by the duration of digital health interventions (DHIs) [34,35,37-43]. REML: restricted maximum likelihood.



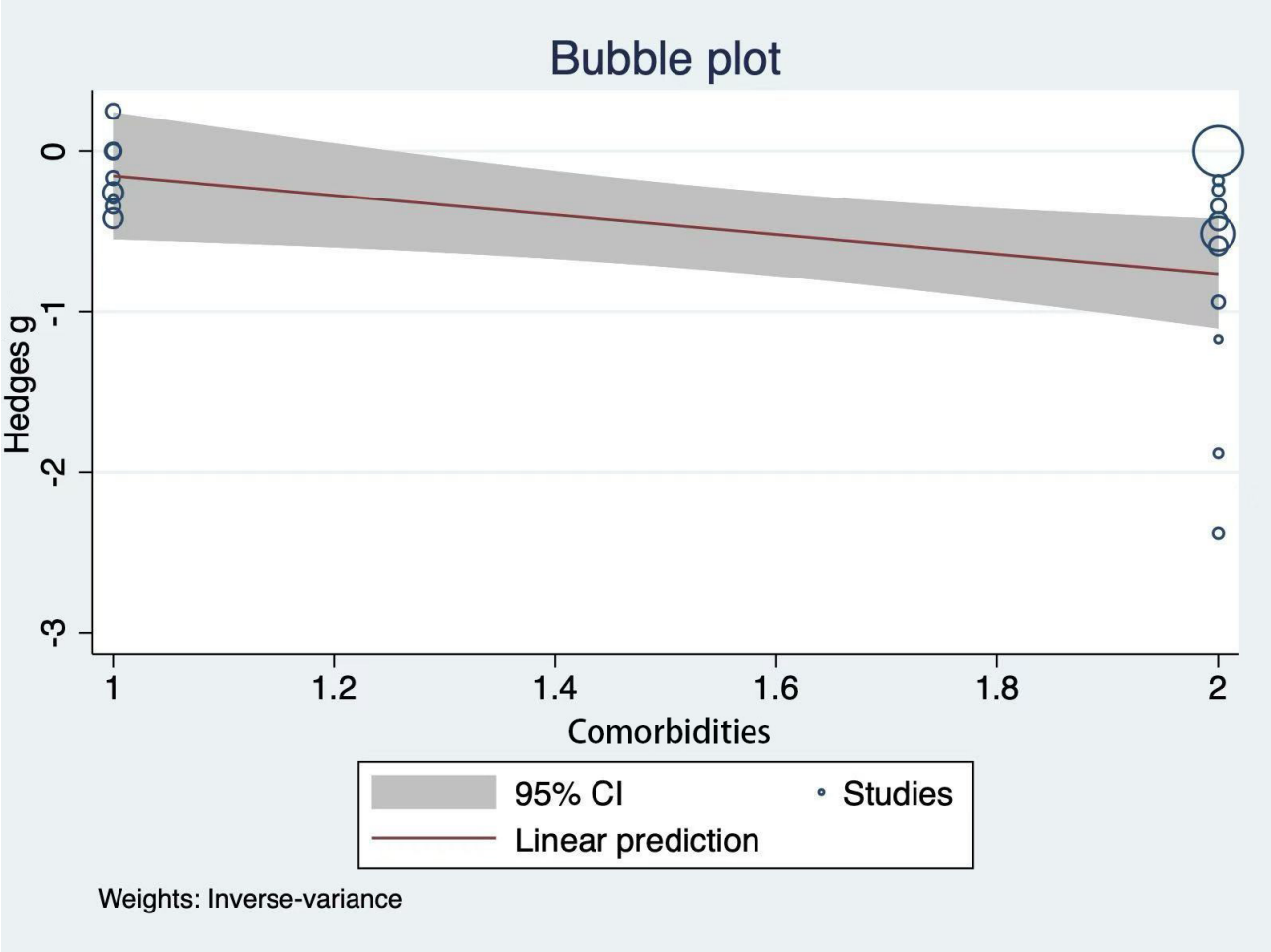
In the subgroup analysis of duration of DHIs in Figure 4, only the 6-month intervention group demonstrated a significant reduction in HbA_{1c} (P<.001). The test for subgroup differences indicates that the effectiveness of DHIs varies significantly across different intervention durations (P<.001). This 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.

Figure 5. Meta-regression bubble plot showing the relationship between participants' comorbidity status and intervention effects on hemoglobin A_{1c} (HbA_{1c}, %).

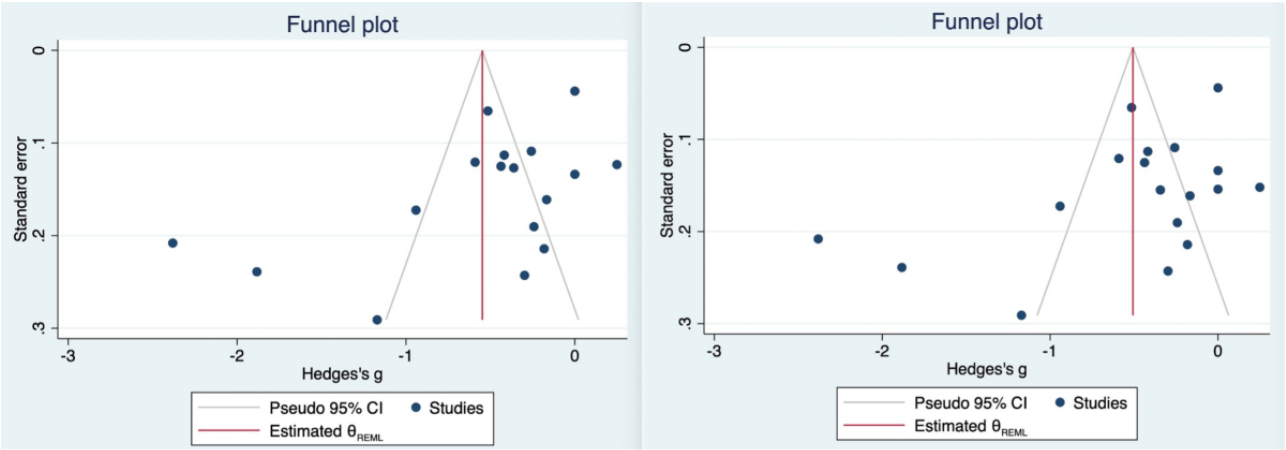


Publication Bias

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.

Figure 6. Publication bias assessment using funnel plots. REML: restricted maximum likelihood.



(a) Original funnel plot with REML estimator

(b) Nonparametric trim-and-fill analysis estimator

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.

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

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

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

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.

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]

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]

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]

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

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

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.

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.

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.

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

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

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,

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

[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.

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.

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]

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]

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

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

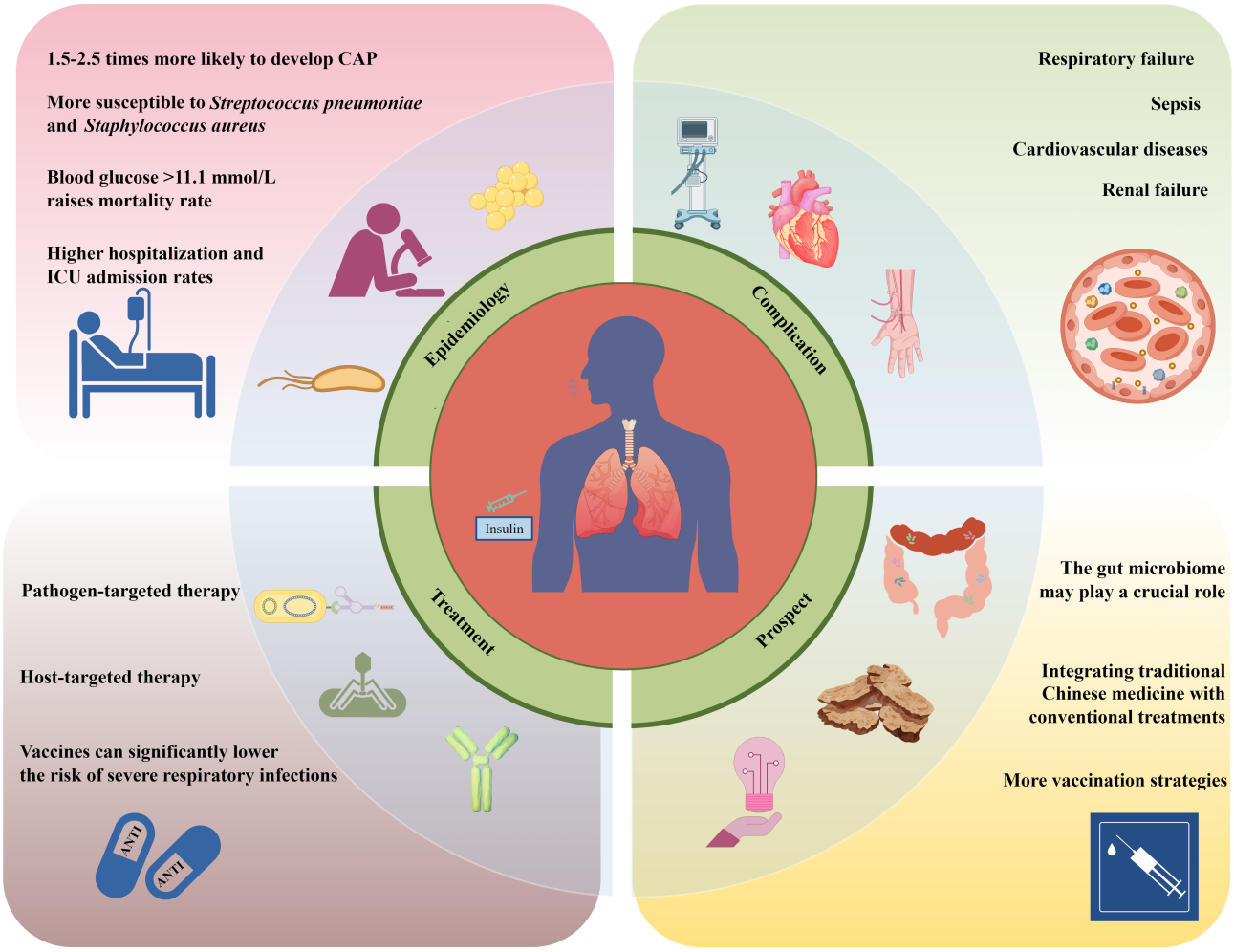
diabetes mellitus; community-acquired pneumonia; pathophysiology; pathogens; complications; treatment; prevention; microbiome

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.

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,

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.

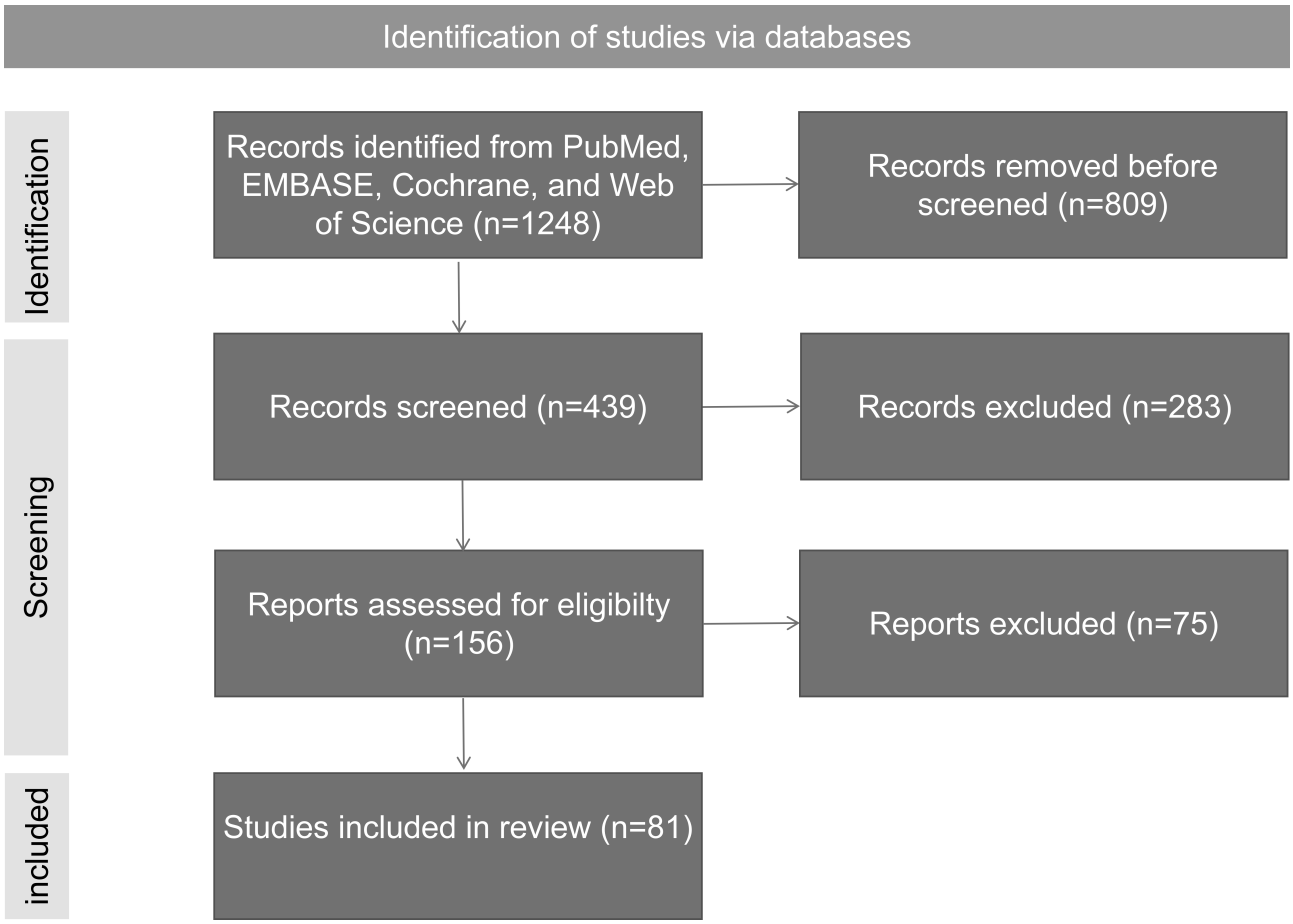


Table . Epidemiological features of community-acquired pneumonia (CAP) in patients with and without diabetes.

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

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).

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.

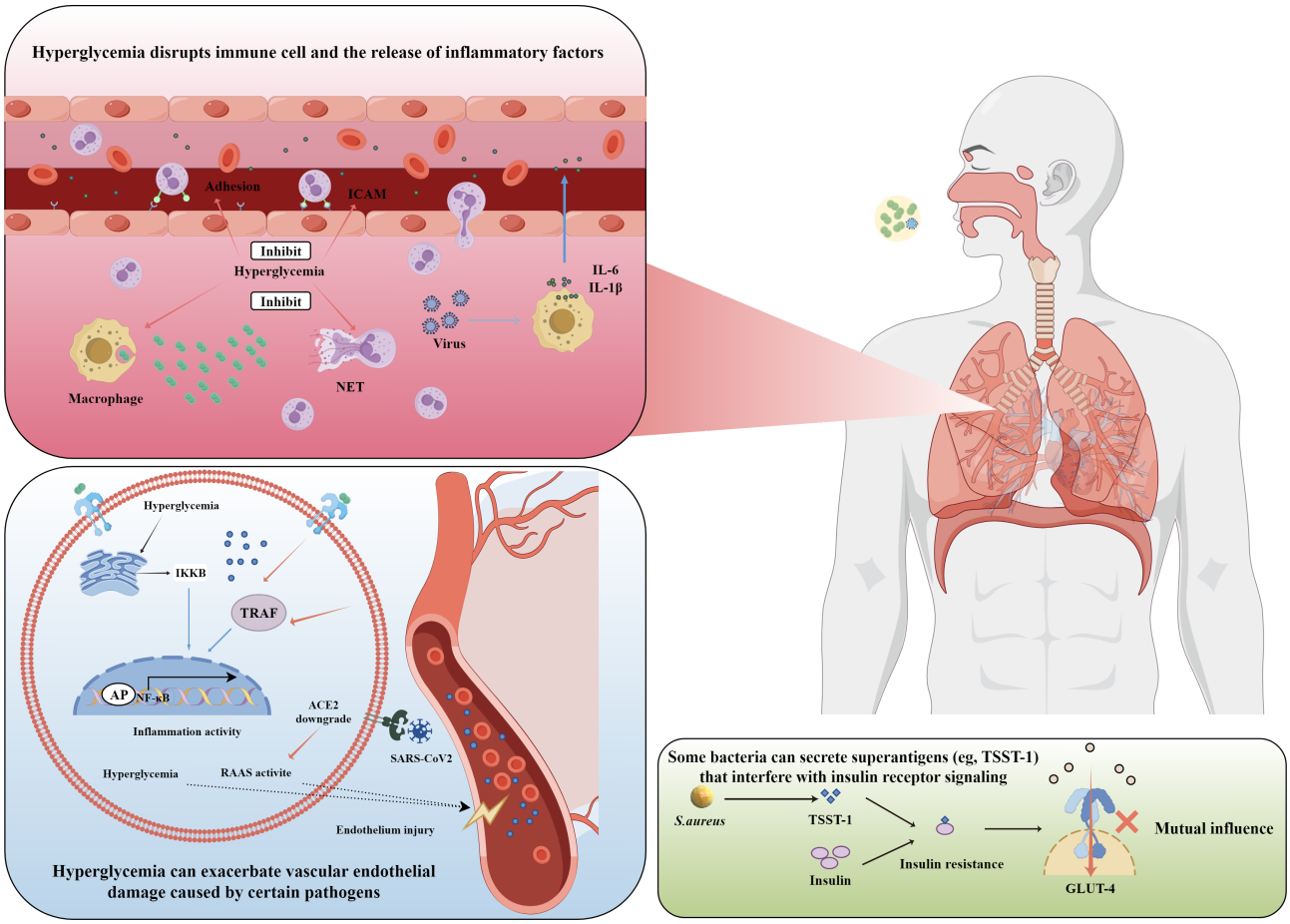


Table . Pathophysiological mechanisms of community-acquired pneumonia (CAP) in patients with diabetes.

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

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).

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.

[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

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.

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]

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]

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]

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

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

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

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 (

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

(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.

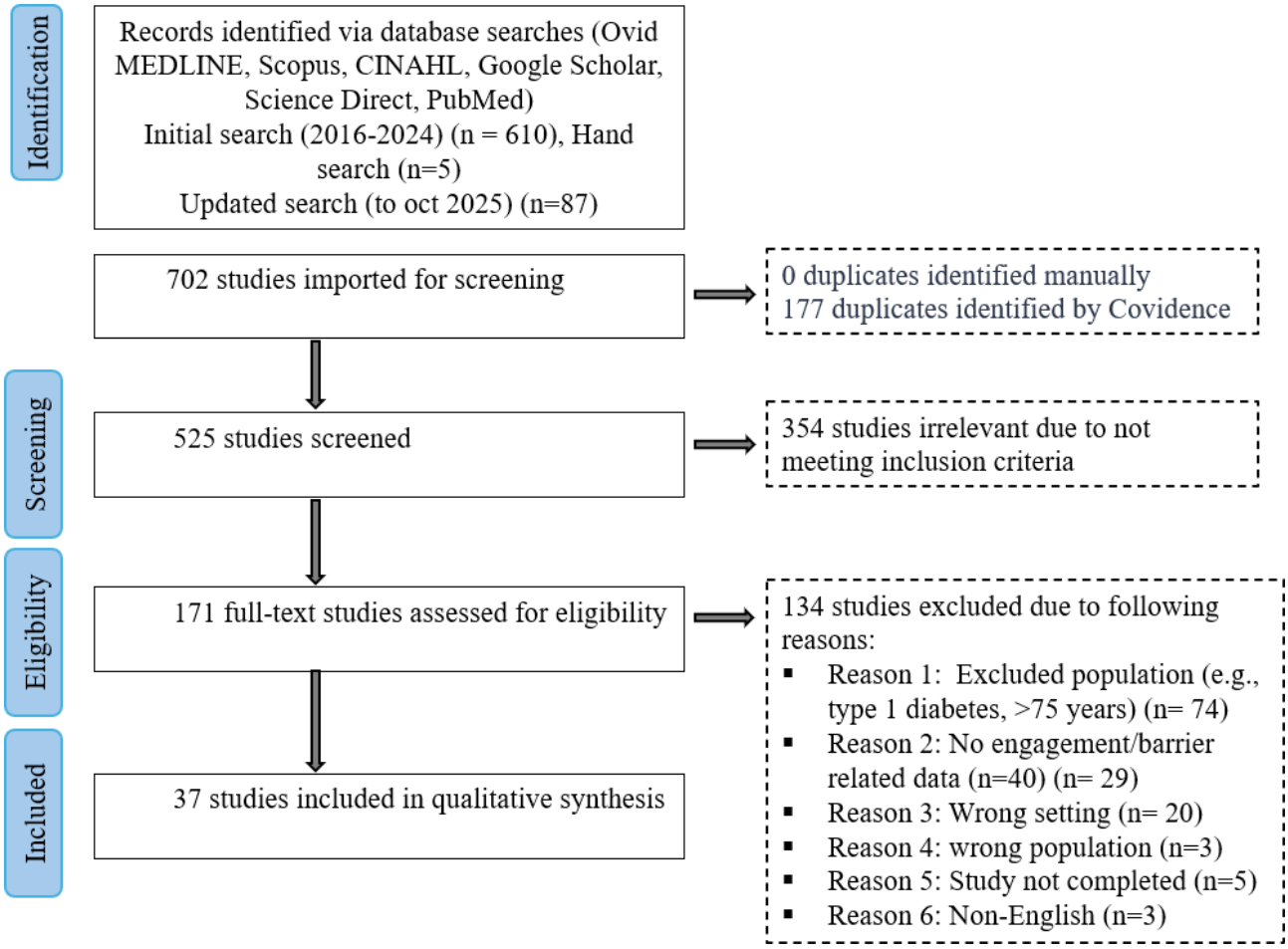
Results

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.

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.

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.

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

[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

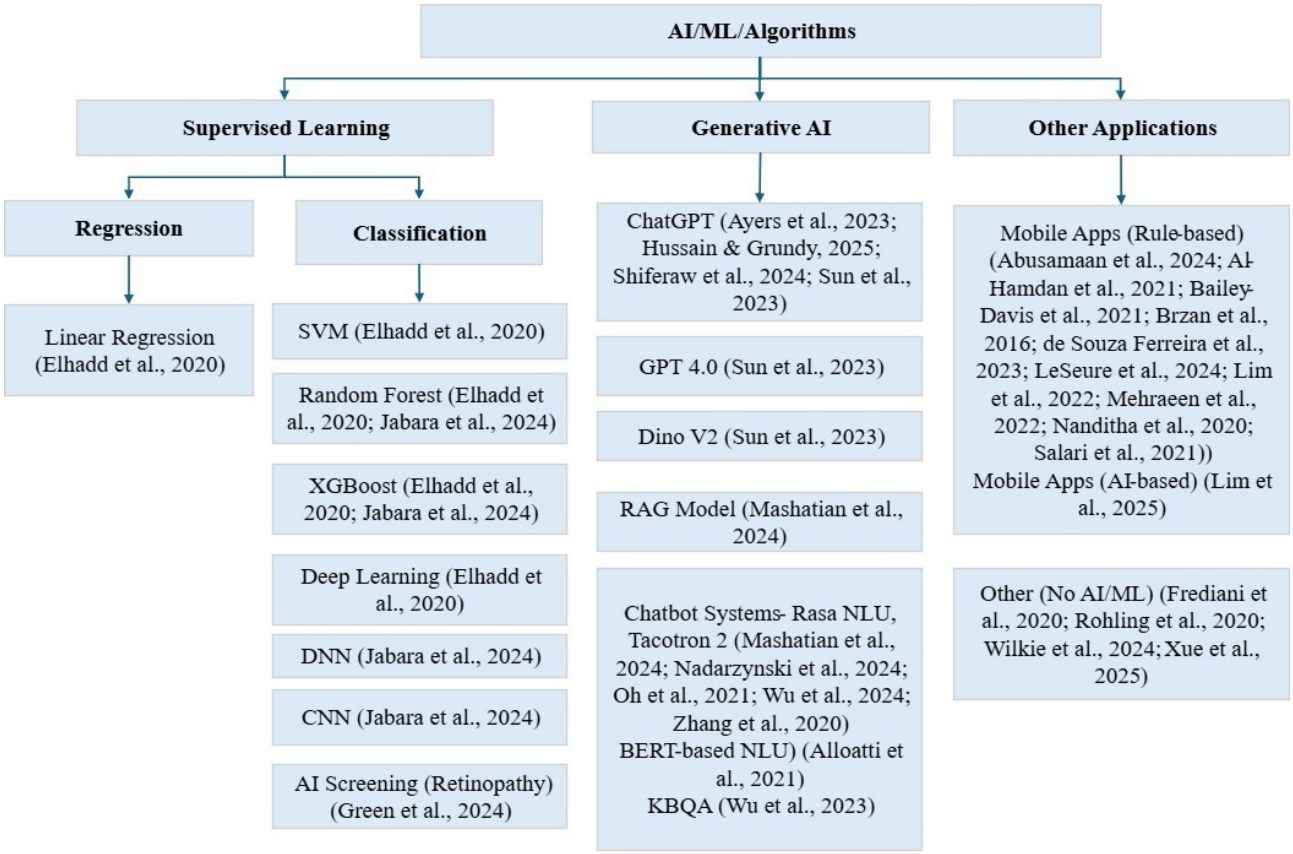
Thematic analysis, guided by SDT and UCD, reveals that AI-based interventions for T2D self-management foster autonomy and competence, yielding significant clinical behavioral outcomes, including HbA_{1c} reductions (0.19% - 1.8%), improved diet and physical activity adherence, and weight loss (0.8% - 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.

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.

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_{1c}[b] 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

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_{1c}[f] 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,

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.

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]

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]

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]

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

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

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].

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

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,

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

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.

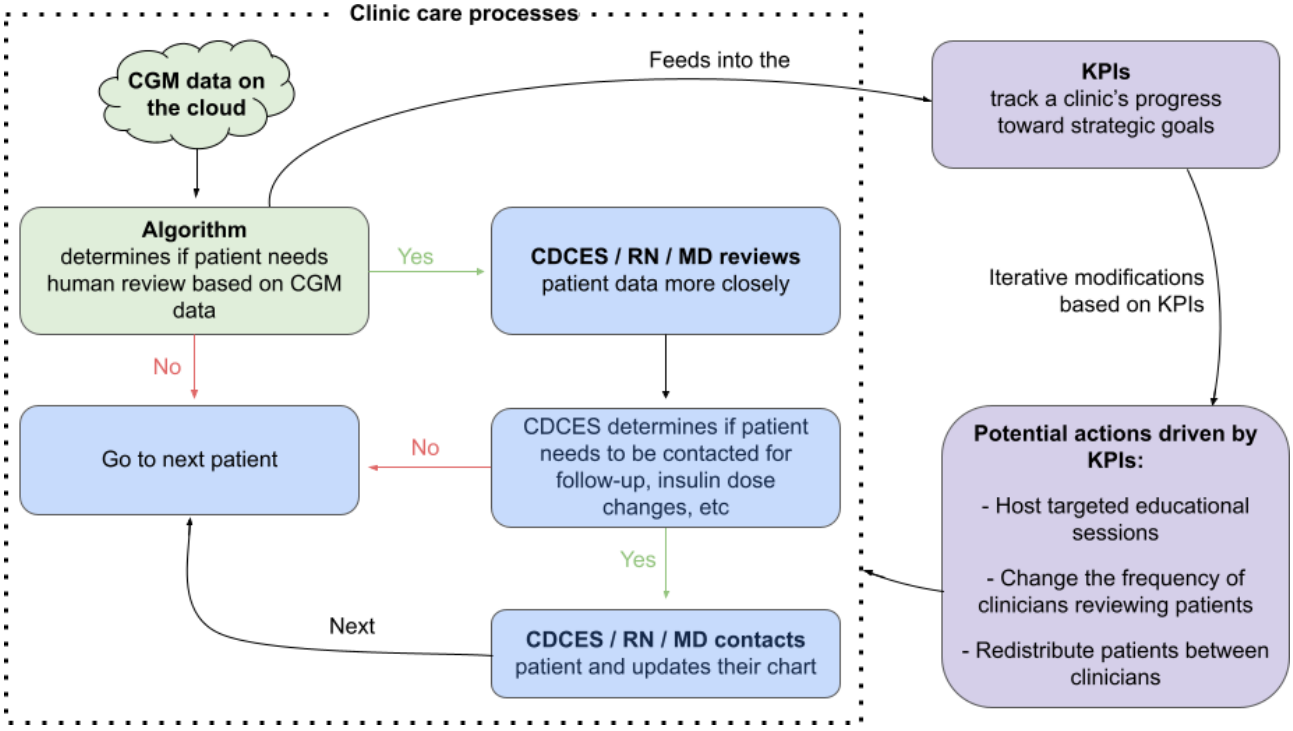
[c]CDCES: Certified Diabetes Care and Education Specialist.

[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.

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.

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)

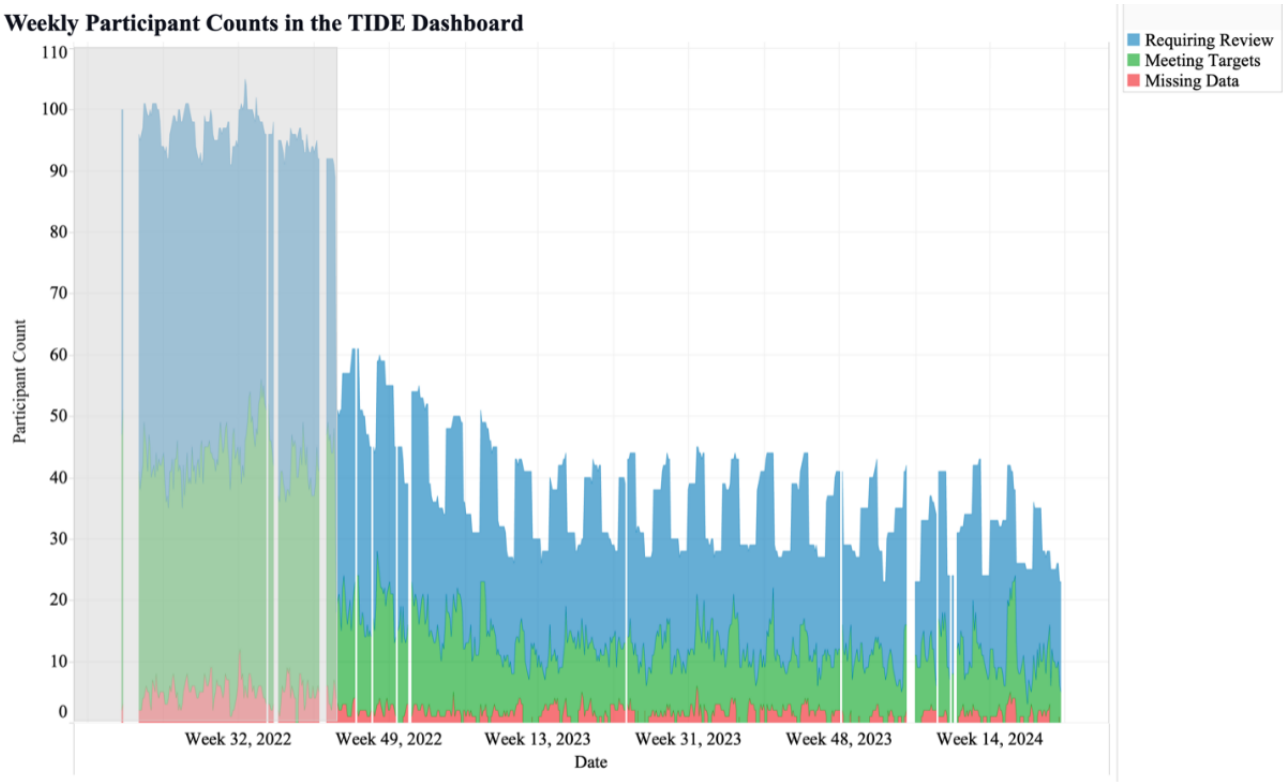
[a]TIDE: Timely Interventions for Diabetes Excellence.

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

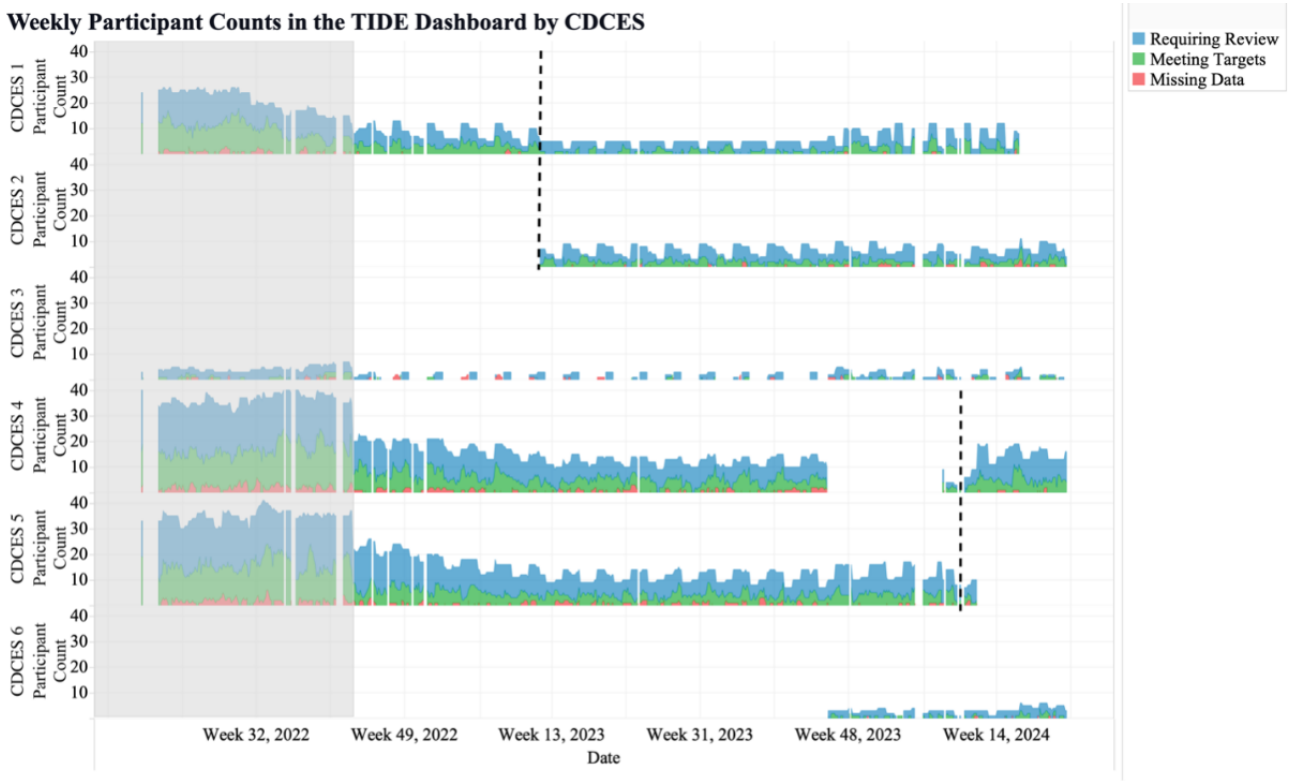
The first metric was the total number of participants shown in the TIDE dashboard. By monitoring the number of participants shown in TIDE, this metric allowed for the tracking of population statistics over time. The visualization of the metric selected by the CDCES team included the total number of participants 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.

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.



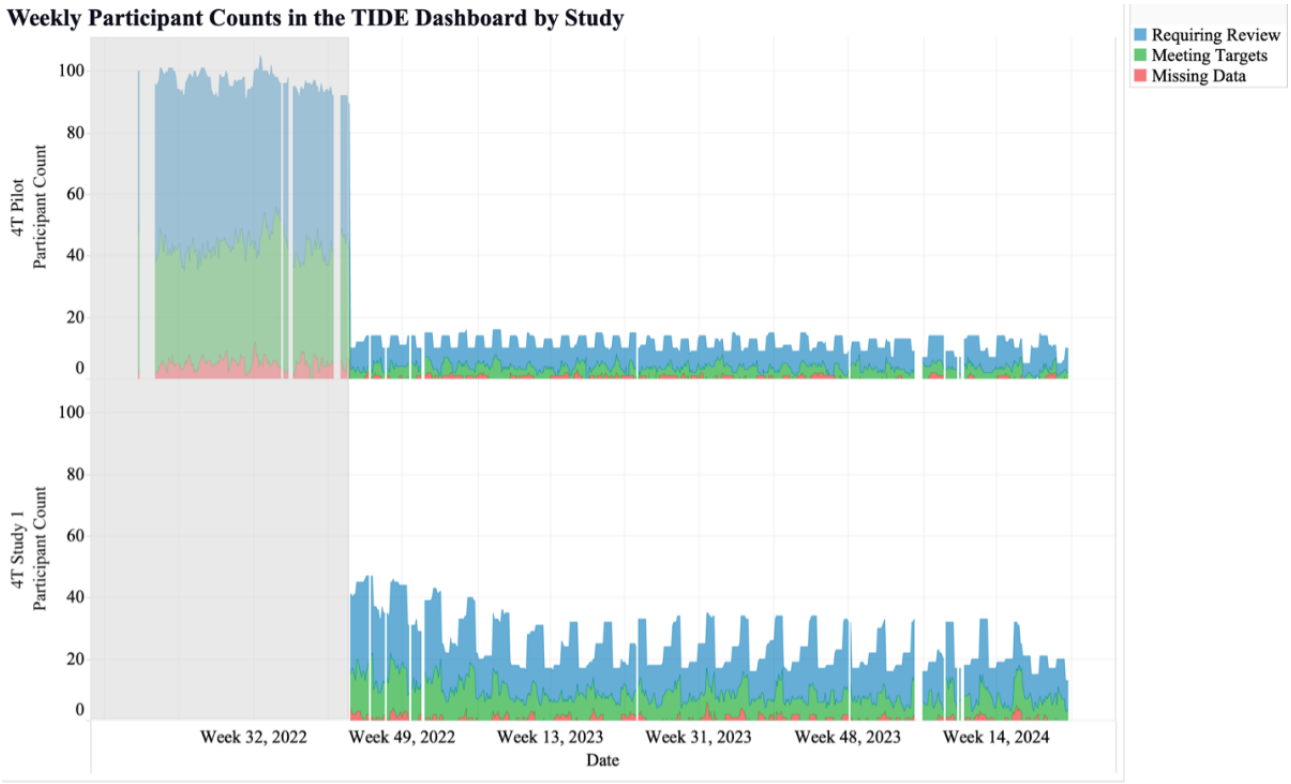
The second metric was the number of participants shown in TIDE, categorized by each CDCES in the program. CDCES names are numbered to protect anonymity. This metric revealed significant differences in the number of participants assigned to each CDCES, as well as stability in the number of participants 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.

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.

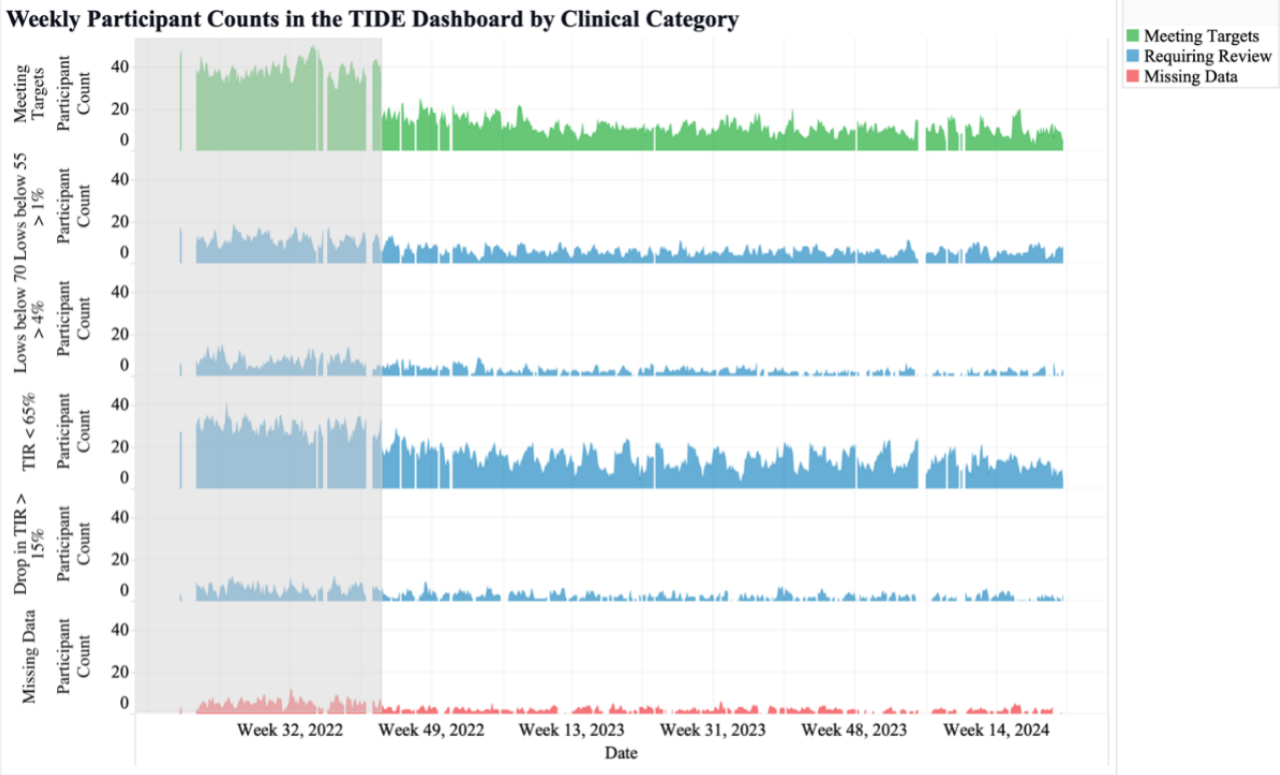
Figure 4. Weekly participant counts per study in the Timely Interventions for Diabetes Excellence dashboard, categorized by participants requiring review (blue), meeting targets (green), and with more than 50% missing data (red). TIDE: Timely Interventions for Diabetes Excellence.



Metrics and Visualizations for Patient Outcomes

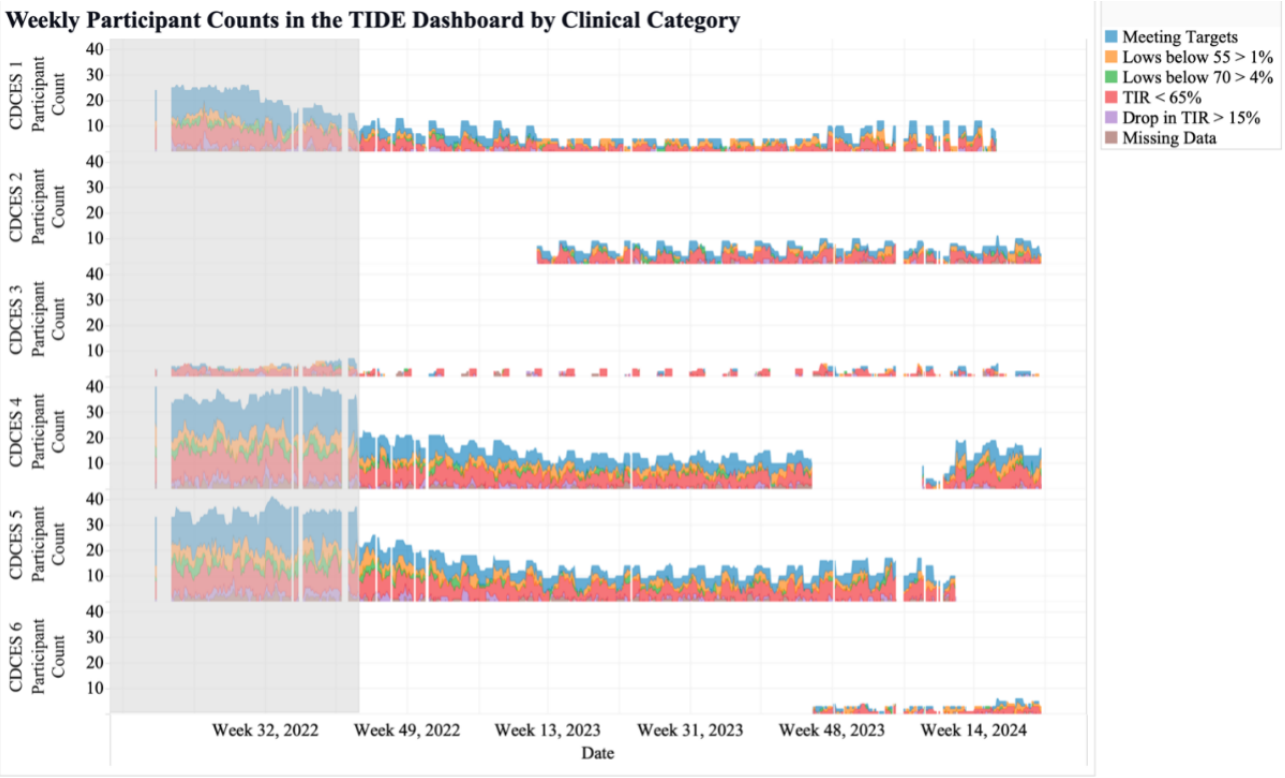
The fourth metric was the number of participants shown in TIDE per clinical category. This metric identified the level of variability in glucose management over time (Figure 5). This metric may help monitor participants' data and identify areas of improvement or concern, allowing CDCES team members to 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.

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.



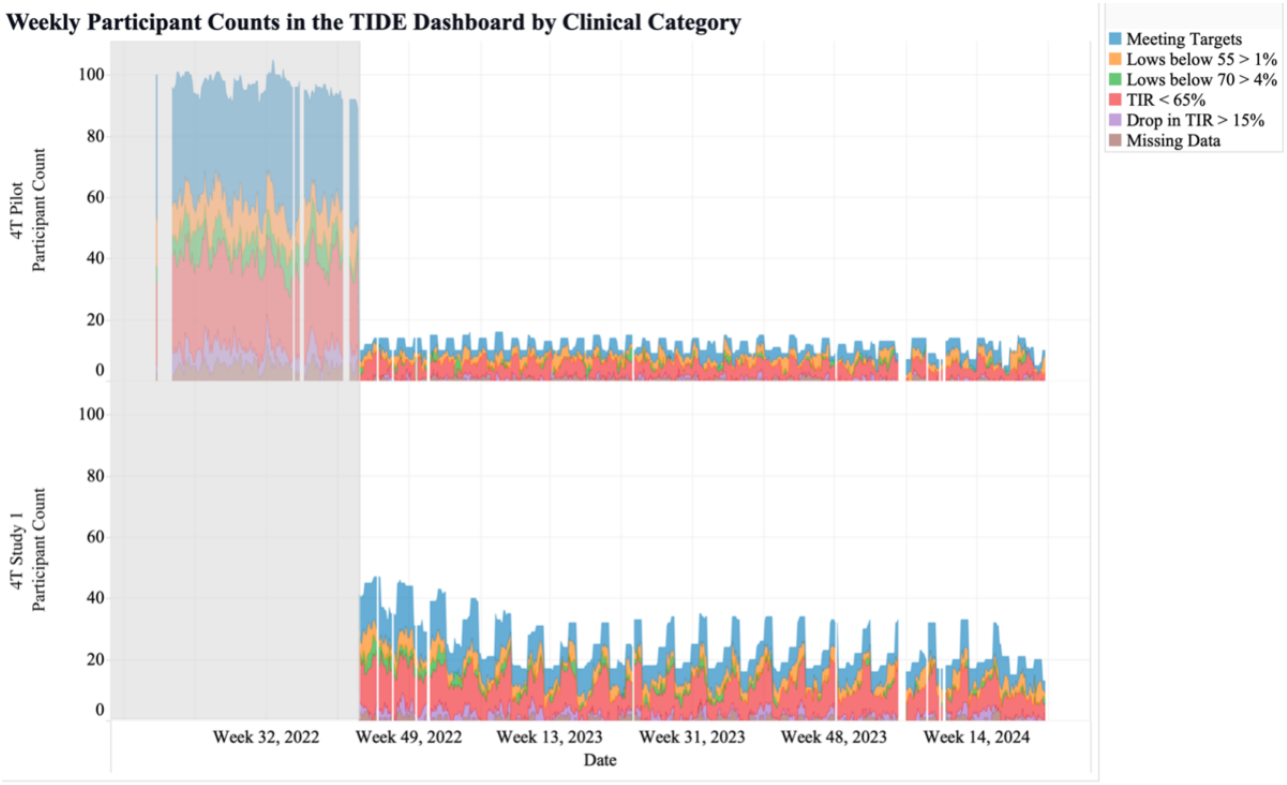
The fifth metric was the number of participants shown in TIDE per clinical category and reviewing CDCES. This metric identified the similarities and differences in glucose management for the participants reviewed by each CDCES (Figure 6). In our clinic, the metric revealed that all CDCES team members 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.

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).

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].

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

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

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]

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

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

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

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.

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

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.

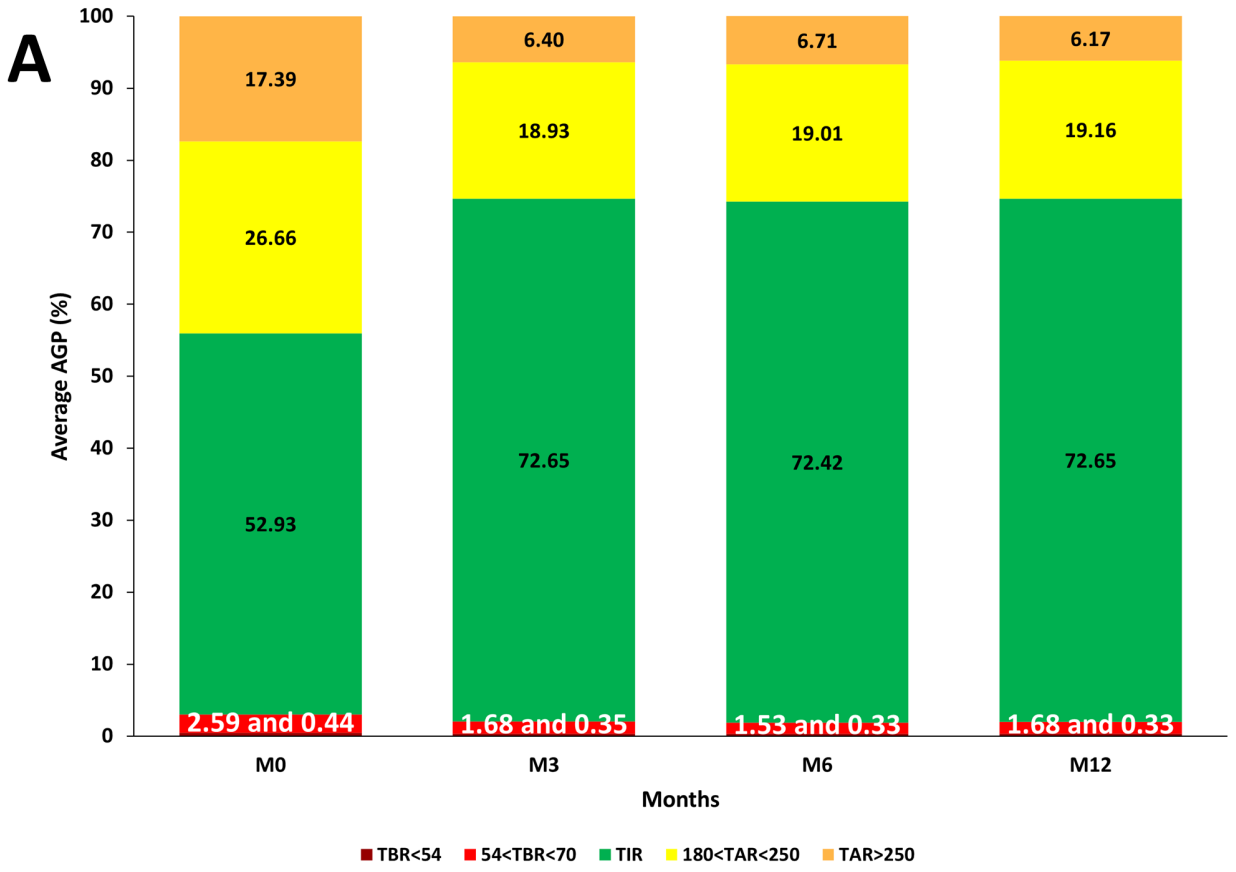
[e]Wilcoxon test values.

[f]GRI: glycemia risk index.

Efficacy and Safety Outcomes

Figure 1 illustrates the changes in the ambulatory glucose profile from baseline to 1 year after initiation.

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.



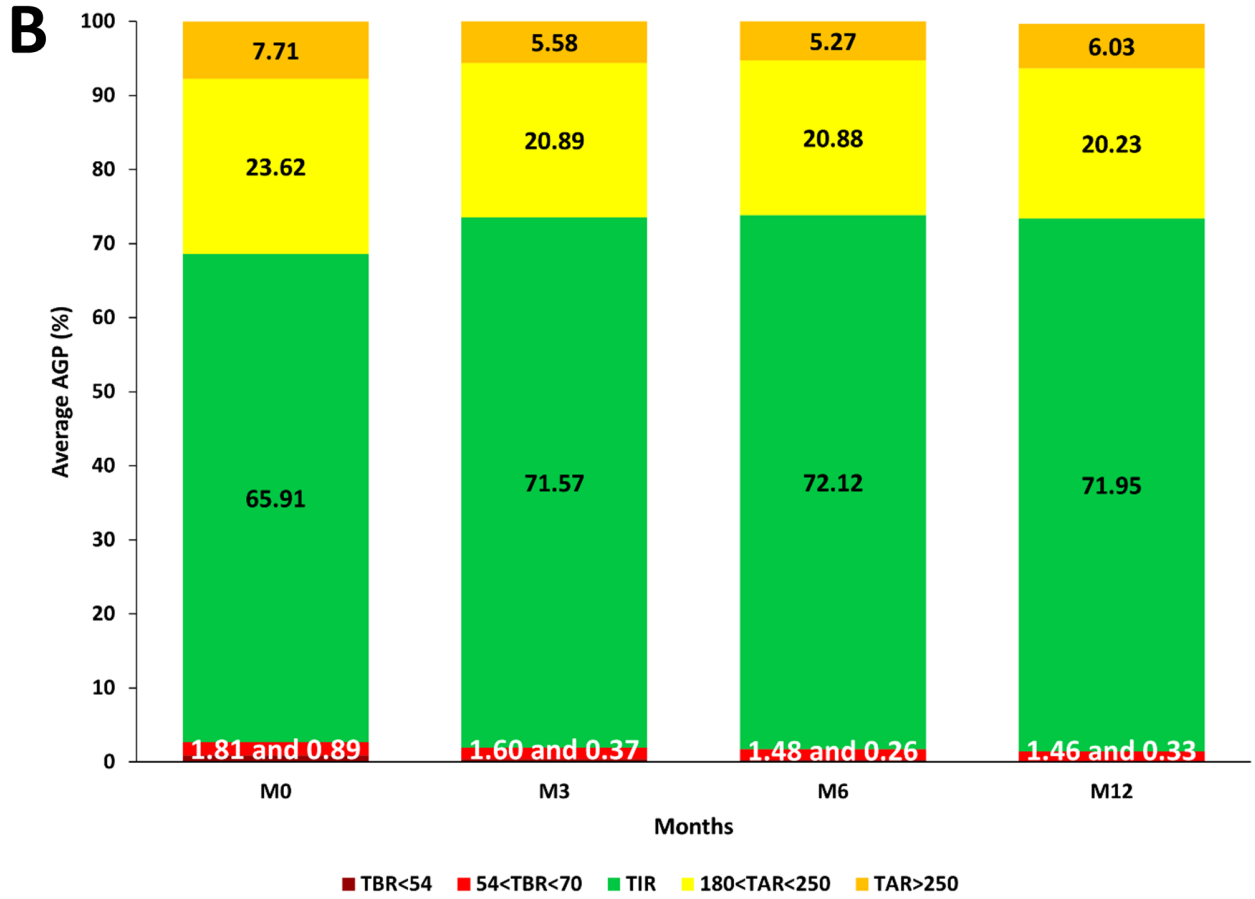


Figure 2 presents GRI grids showing glycemic risk zones over the same period (zone A: minimal hypo- or hyperglycemia risk; zone E: maximal hypo- or hyperglycemia risk). At baseline, 79% (101/128) of the patients from the CIRDIA group were in the intermediate risk (zone C) or high-risk zones (zones D 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.

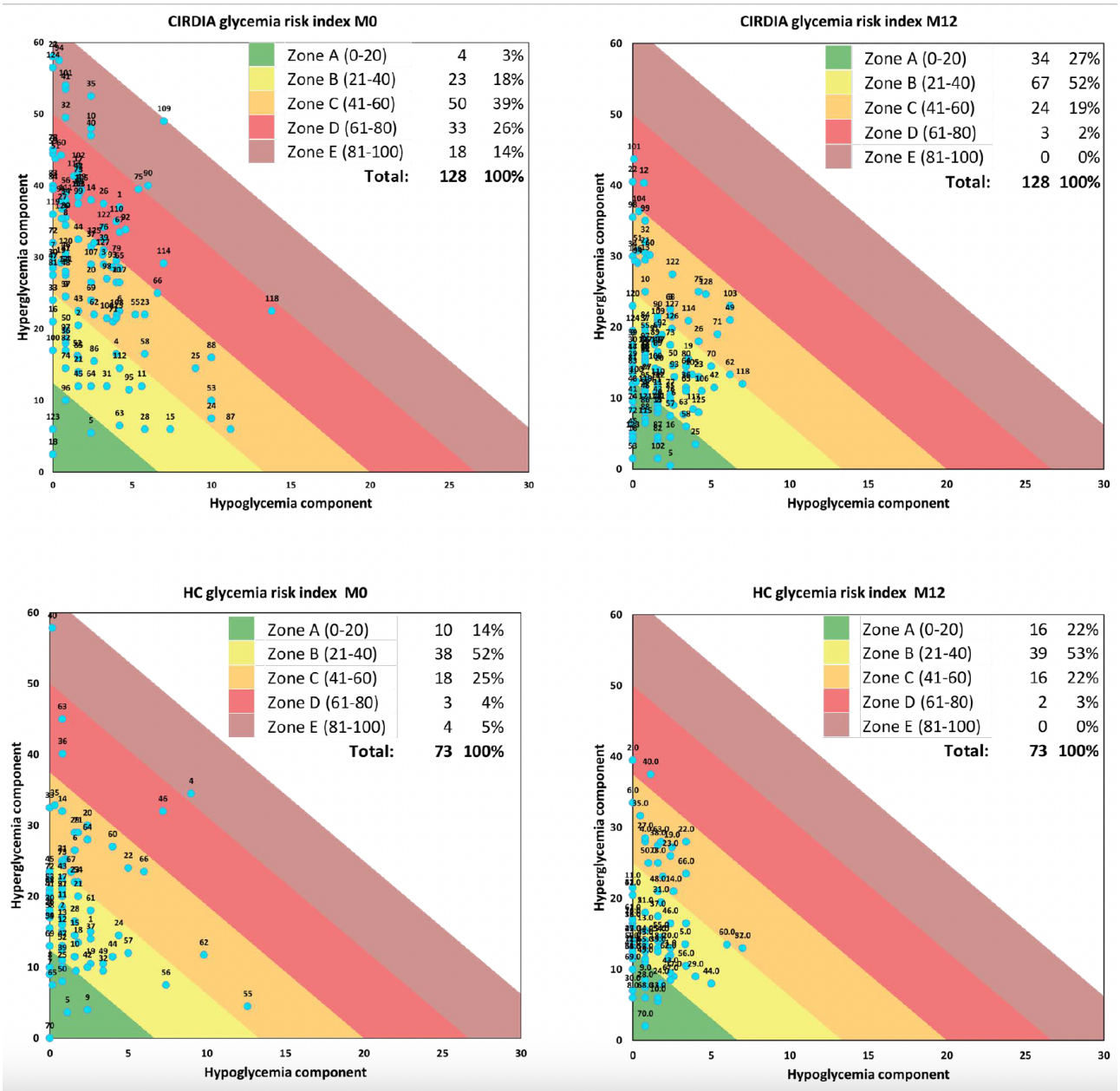


Table 2 summarizes the overall effectiveness and safety results for the total population and according to age at inclusion. After 12 months, the mean (SD) TIR increased by 19.8 points in the CIRDIA group (from 52.9% [16] to 72.7% [11.6]) and by 6 points in the HC group (from 65.9% [15.1] to 71.9% [10.5]). 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).

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.

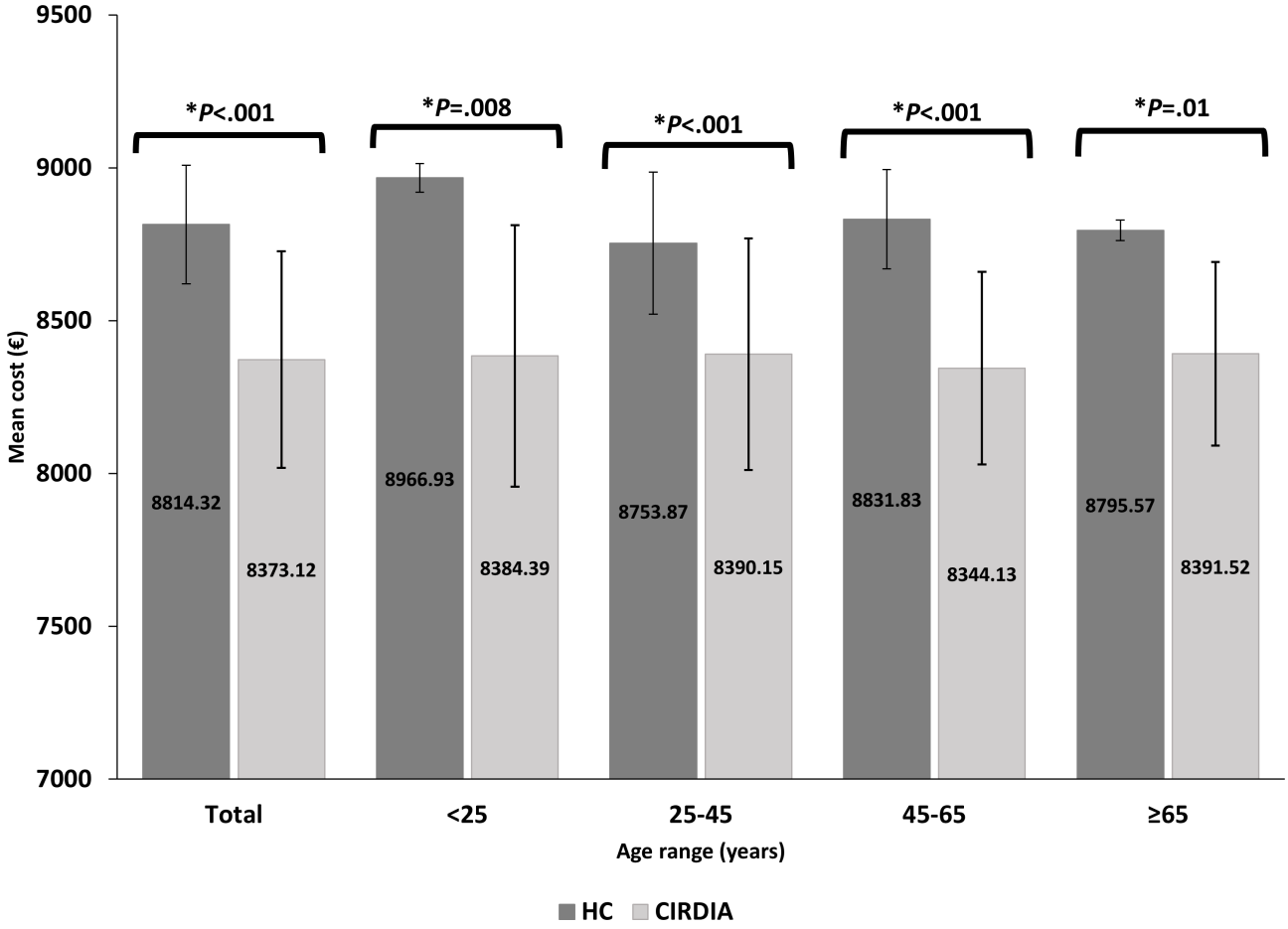
[g]M12: 12 months after closed-loop initiation.

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).

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.

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.

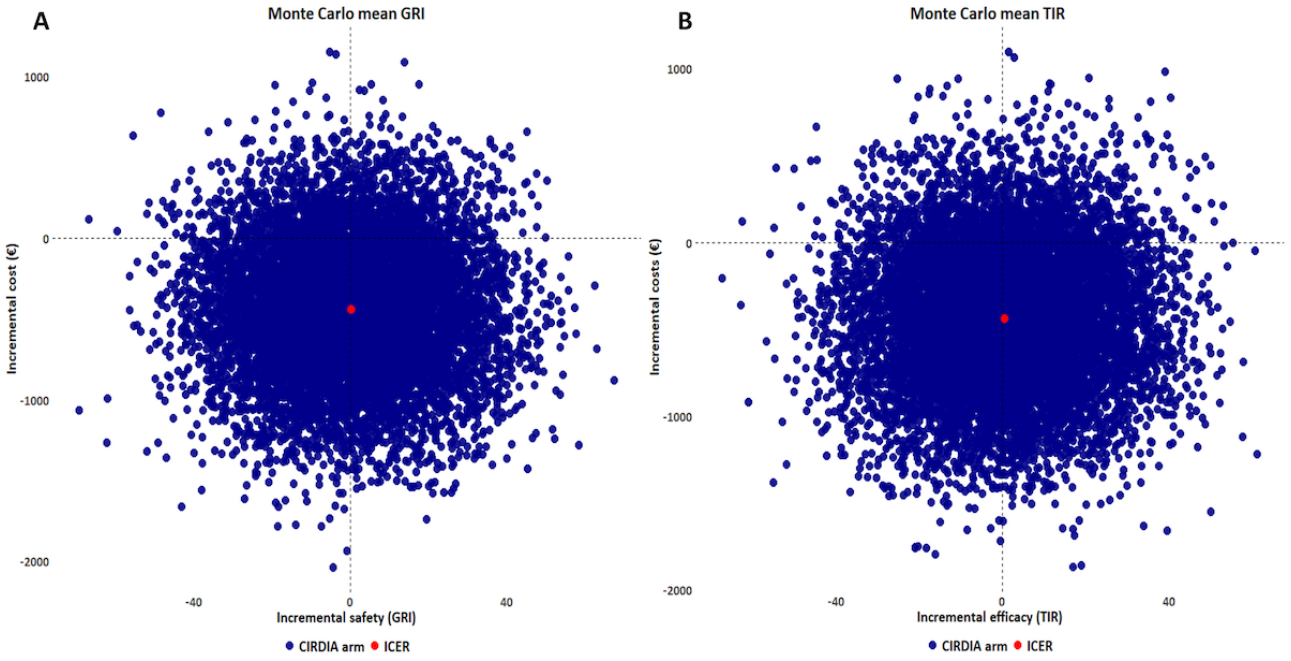
[g]ICER: incremental cost-effectiveness ratio (based on the increase of time in range).

[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.



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

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]

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]

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

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

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.

Figure 1. Example of the GTT@home device including user instructions. GTT: glucose tolerance test.

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

Figure 2. Bias of the GTT@home device. The blue circles indicate fasting levels, while the orange circles indicate the levels after 2 hours.

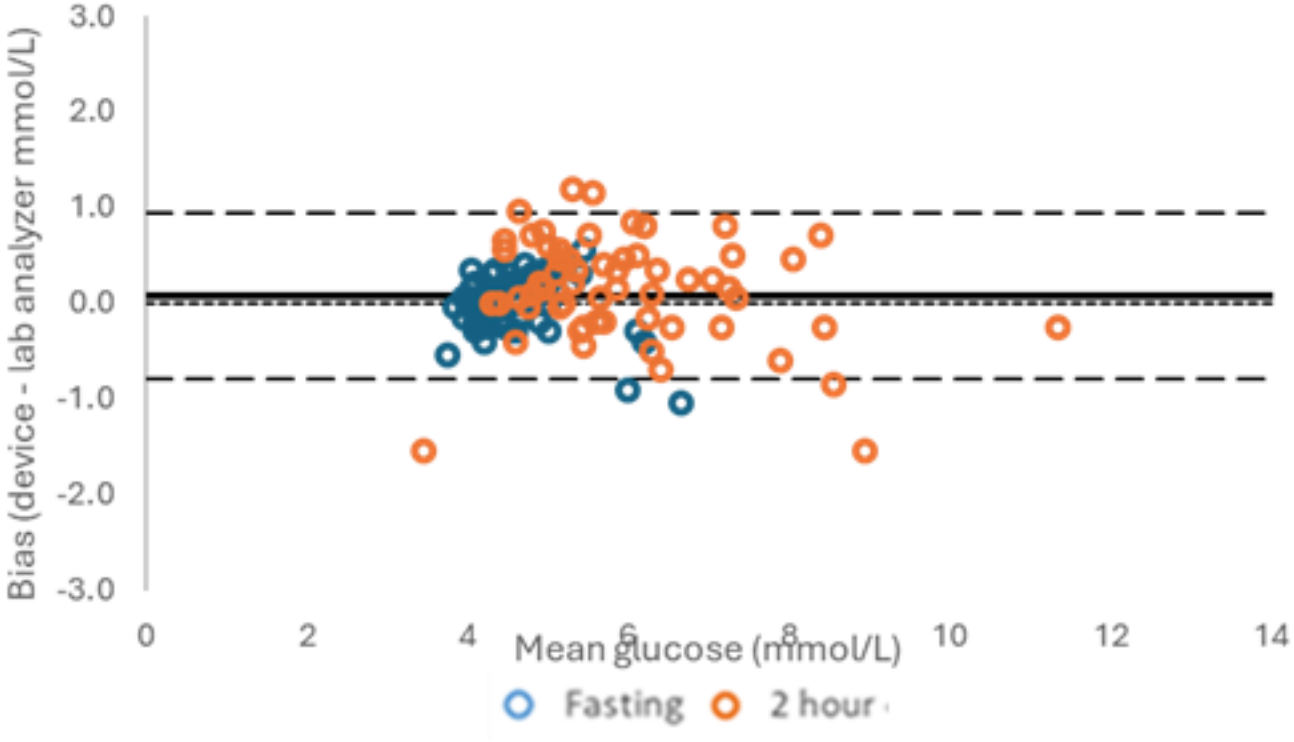
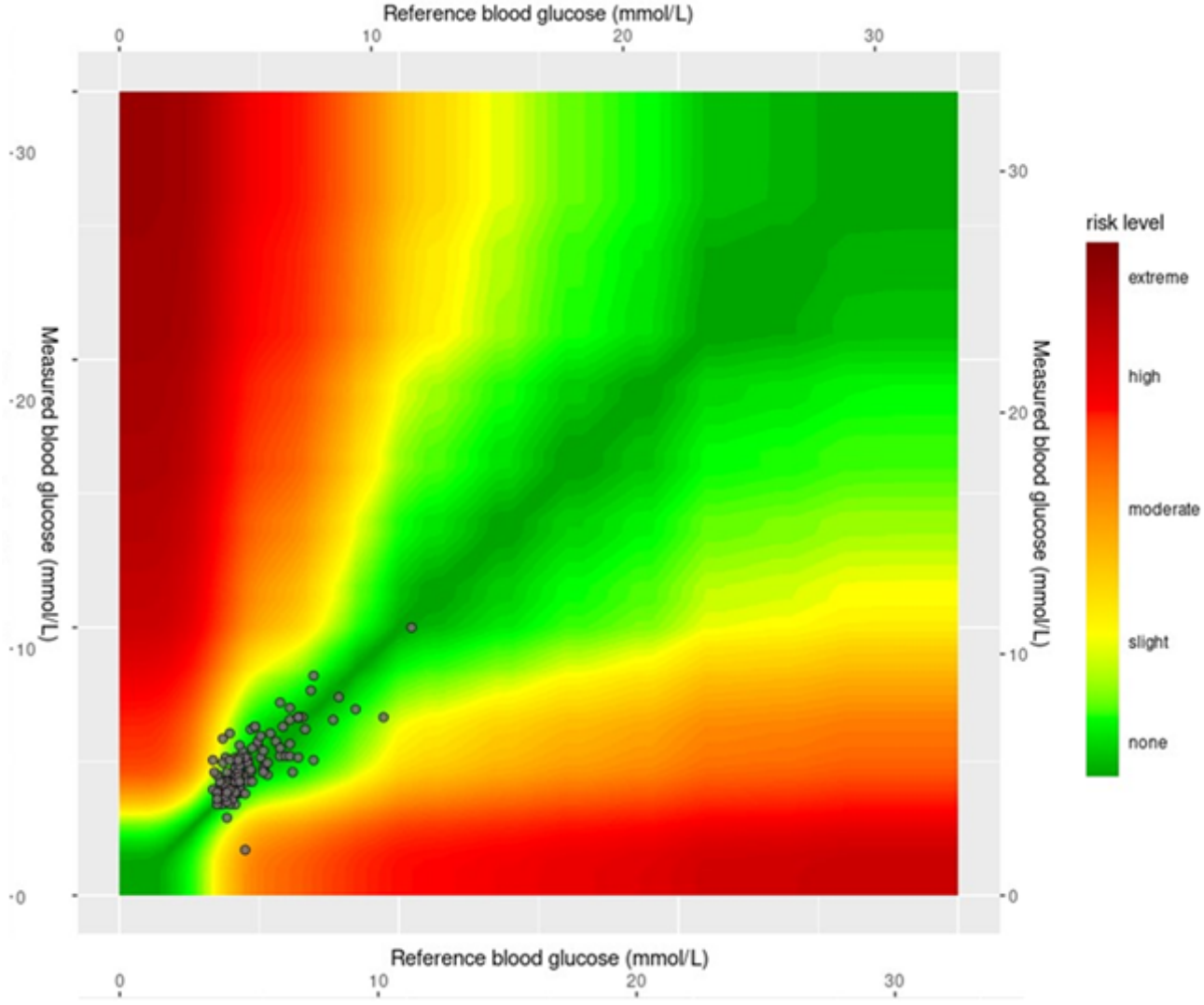


Figure 3. Surveillance error grid (SEG).



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

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.

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.

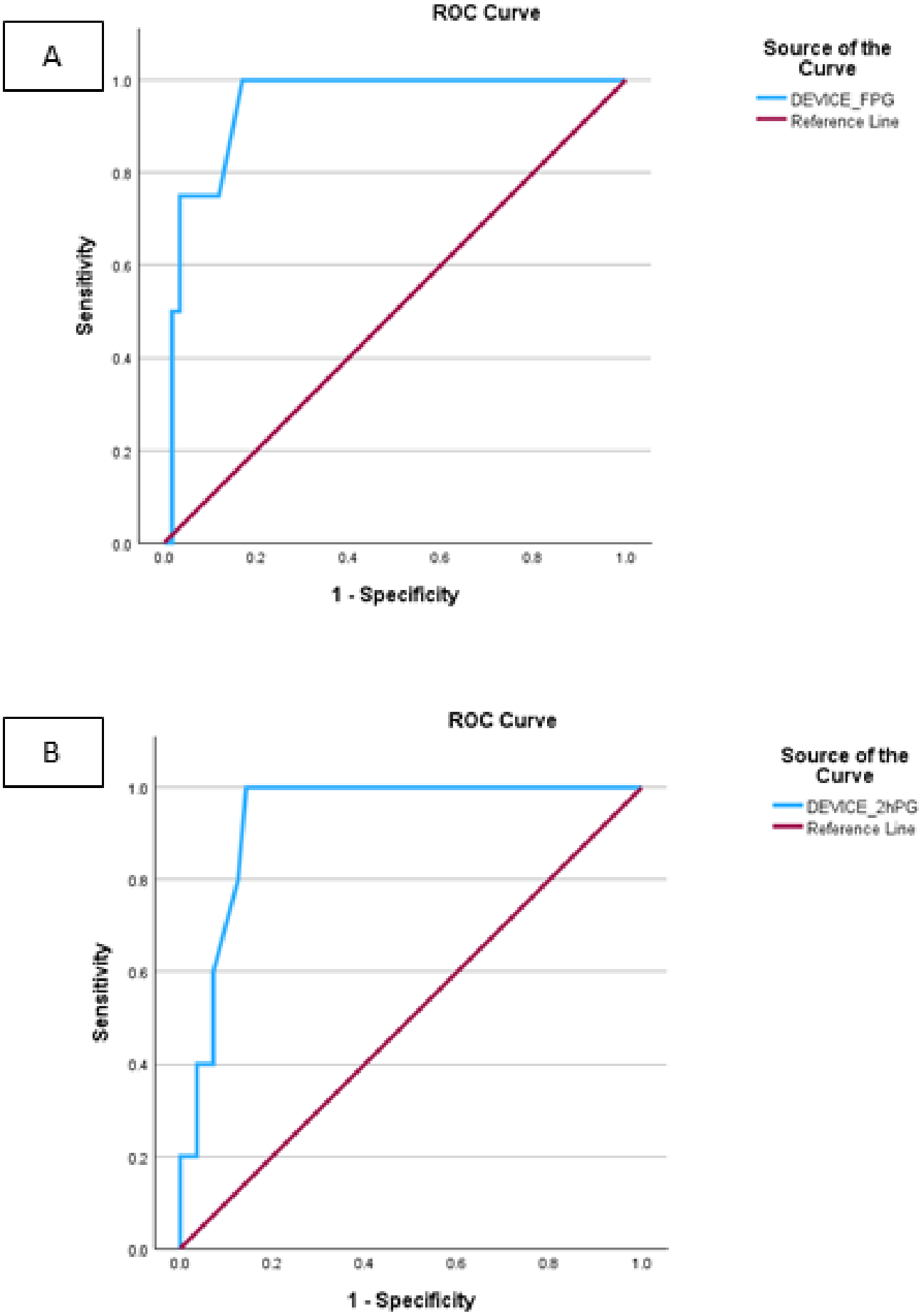


Table . Positive and negative predictive values.

	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.

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

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

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

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

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).

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).

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]

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*

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.

> *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

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.

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]

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]

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]

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

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

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,

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

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

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).

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.

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.

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,

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.

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.

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.

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.

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

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

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

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].

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]

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]

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

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

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.

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.

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).

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)

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.

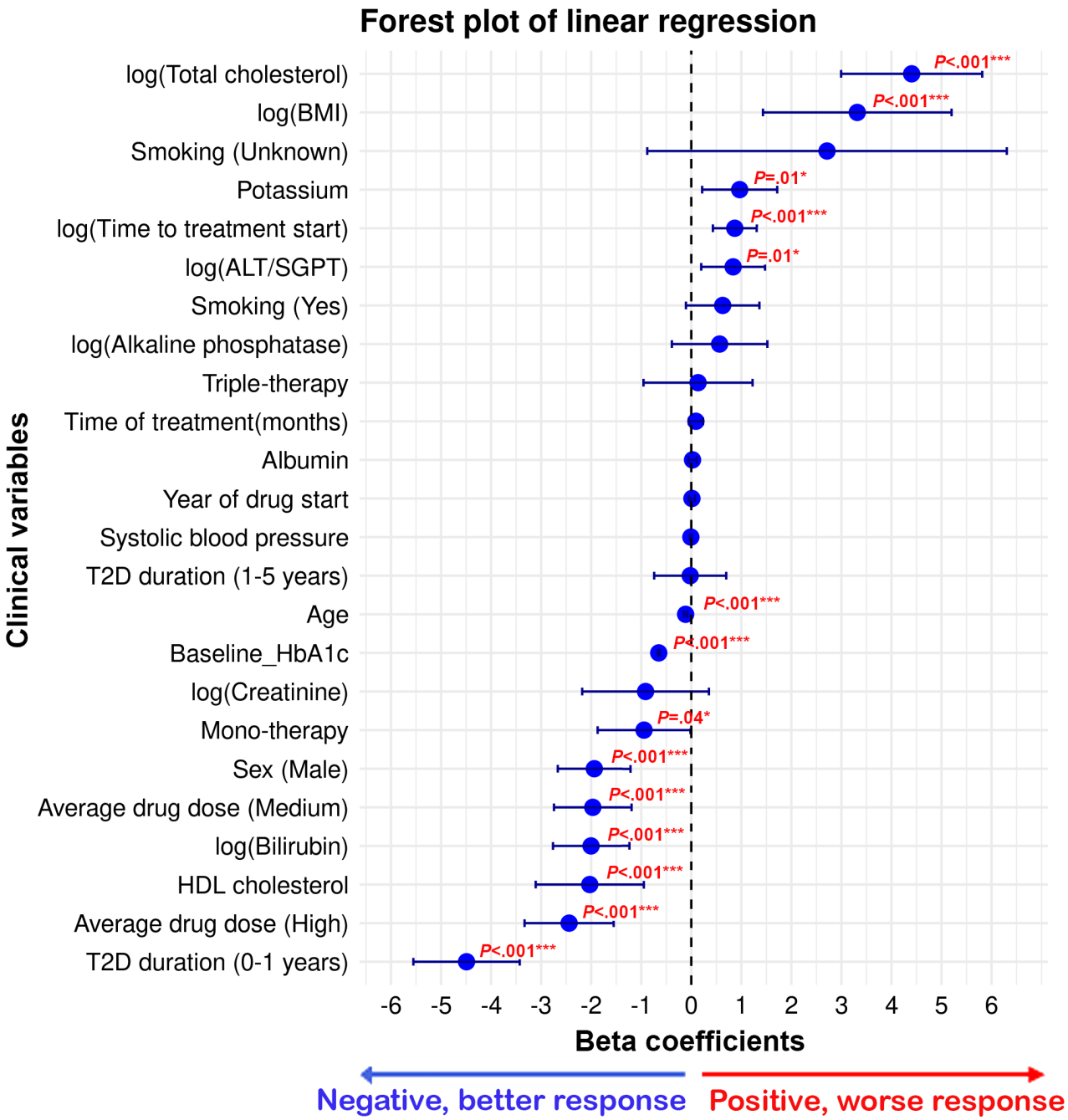
Associations Between Clinical Covariates and Treatment Response

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.

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: β=−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

(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.

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.

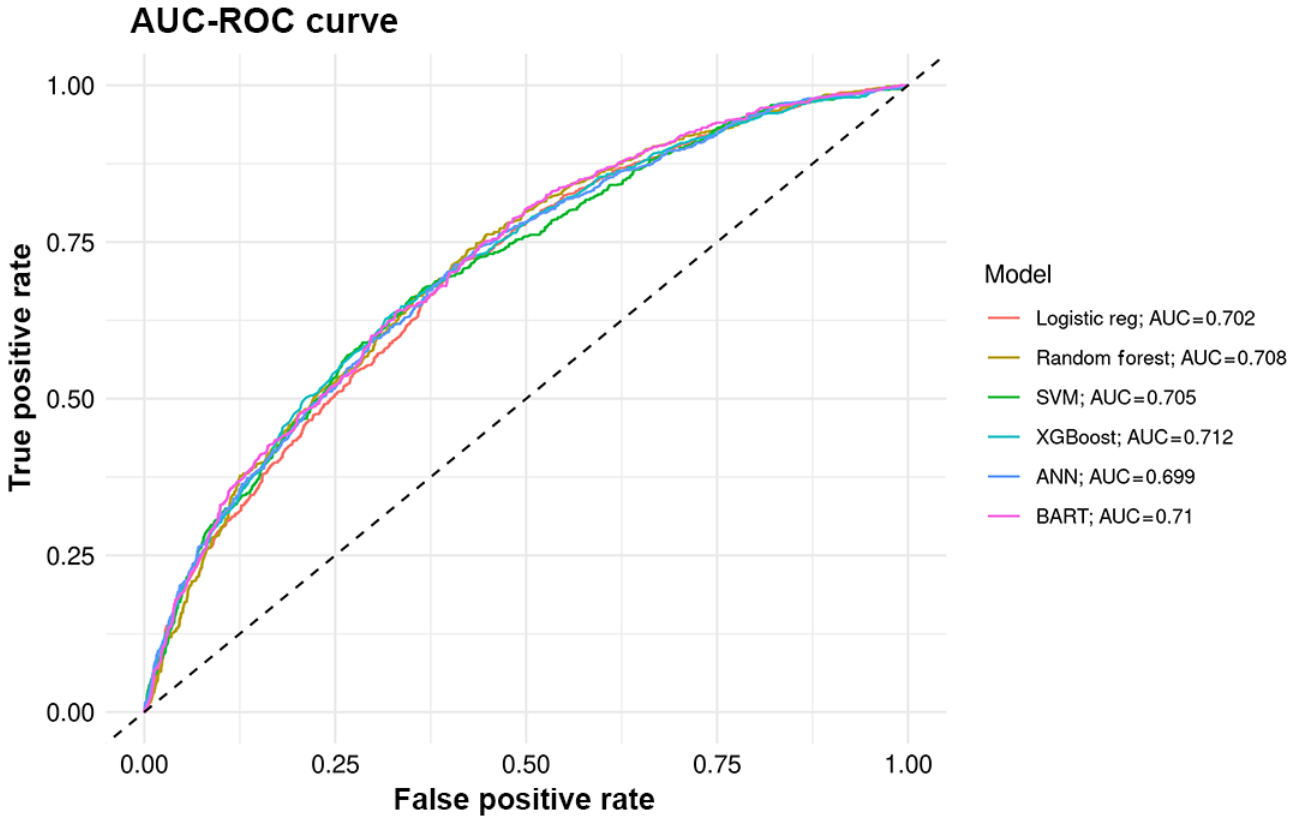
[c]SVM: support vector machine.

[d]XGBoost: extreme gradient boosting.

[e]NN: neural network.

[f]BART: Bayesian additive regression trees.

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.

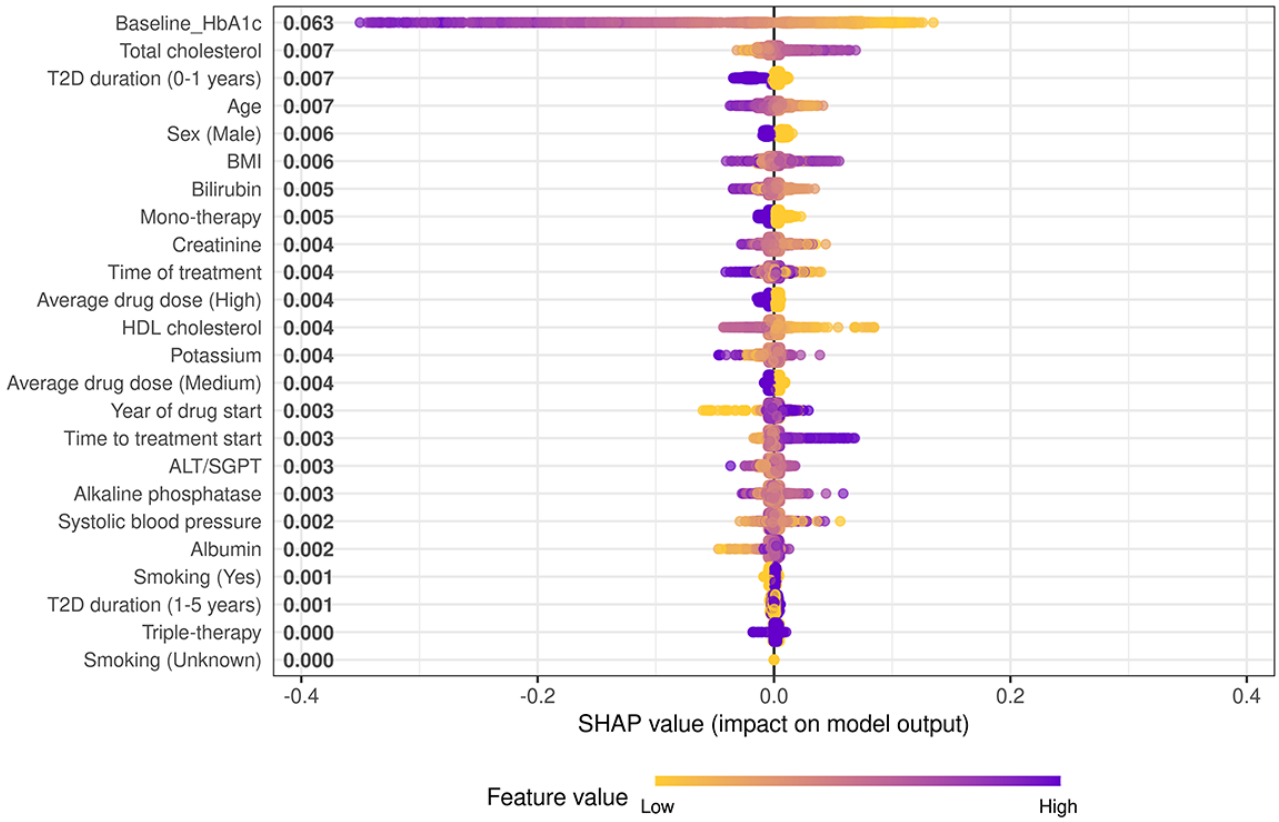


Feature Importance and SHAP Interpretability

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.

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.

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.

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]

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

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

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

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

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.

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.

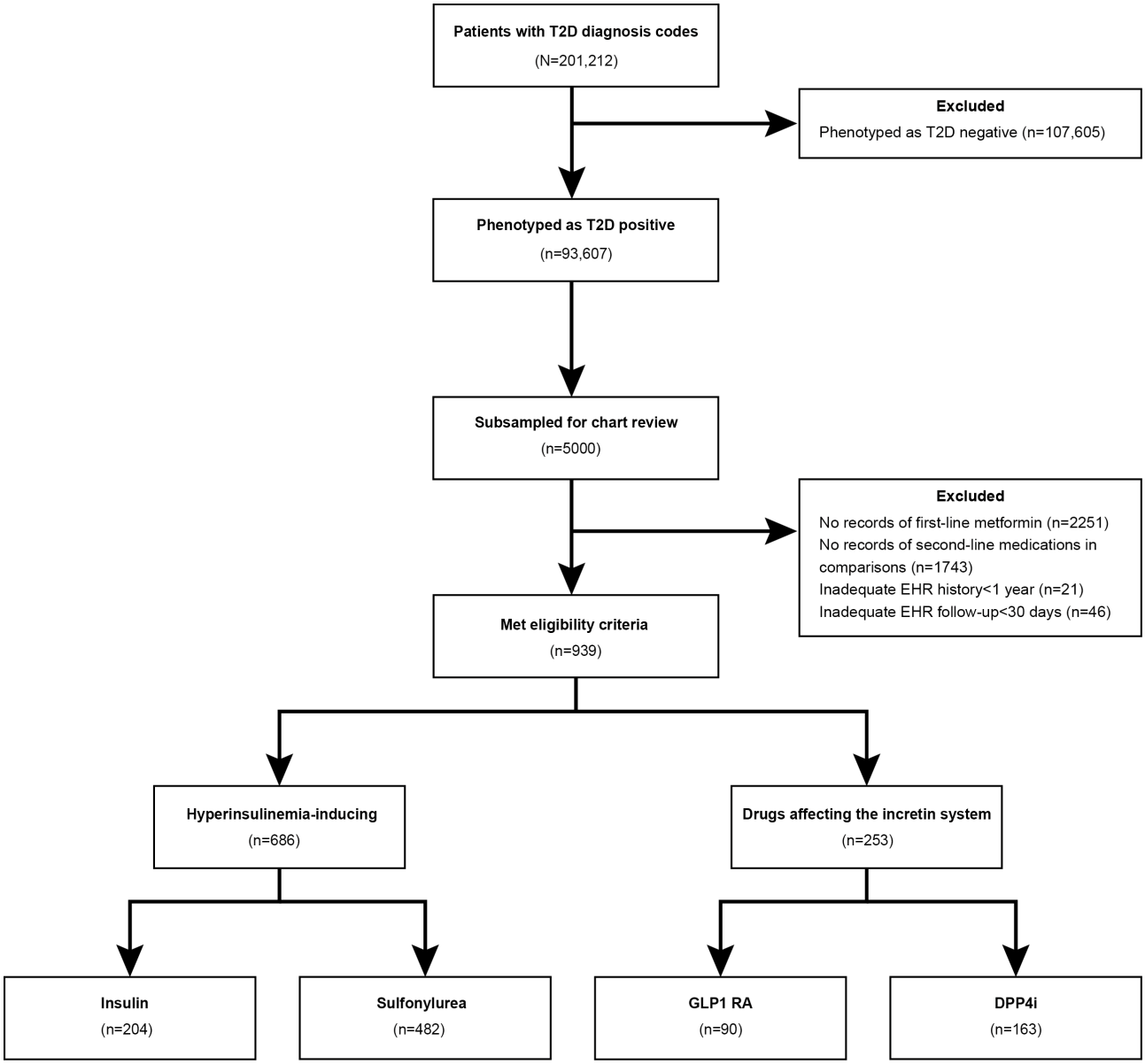


Table . Characteristics of eligible patients included in the study (N=939).

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).

Figure 2. Adjusted confounding in analysis. EHR: electronic health record; SGLT2i: sodium-glucose cotransporter 2 inhibitors.

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.

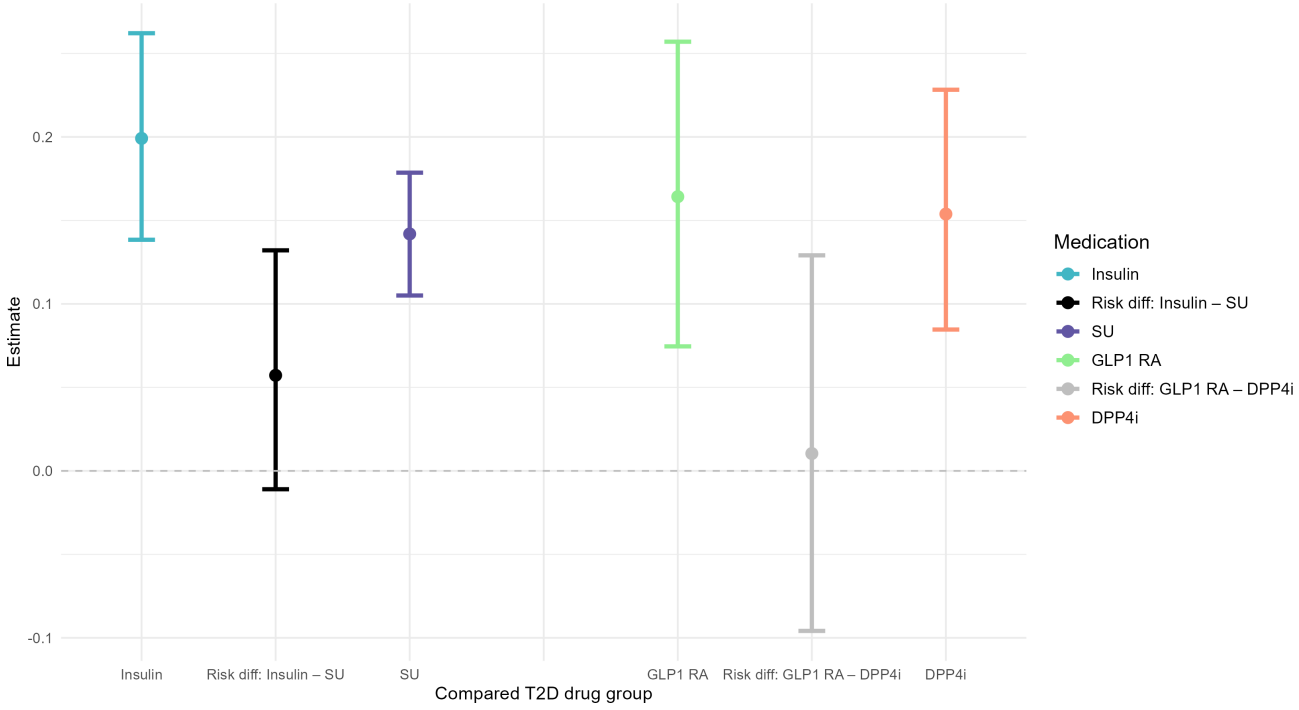
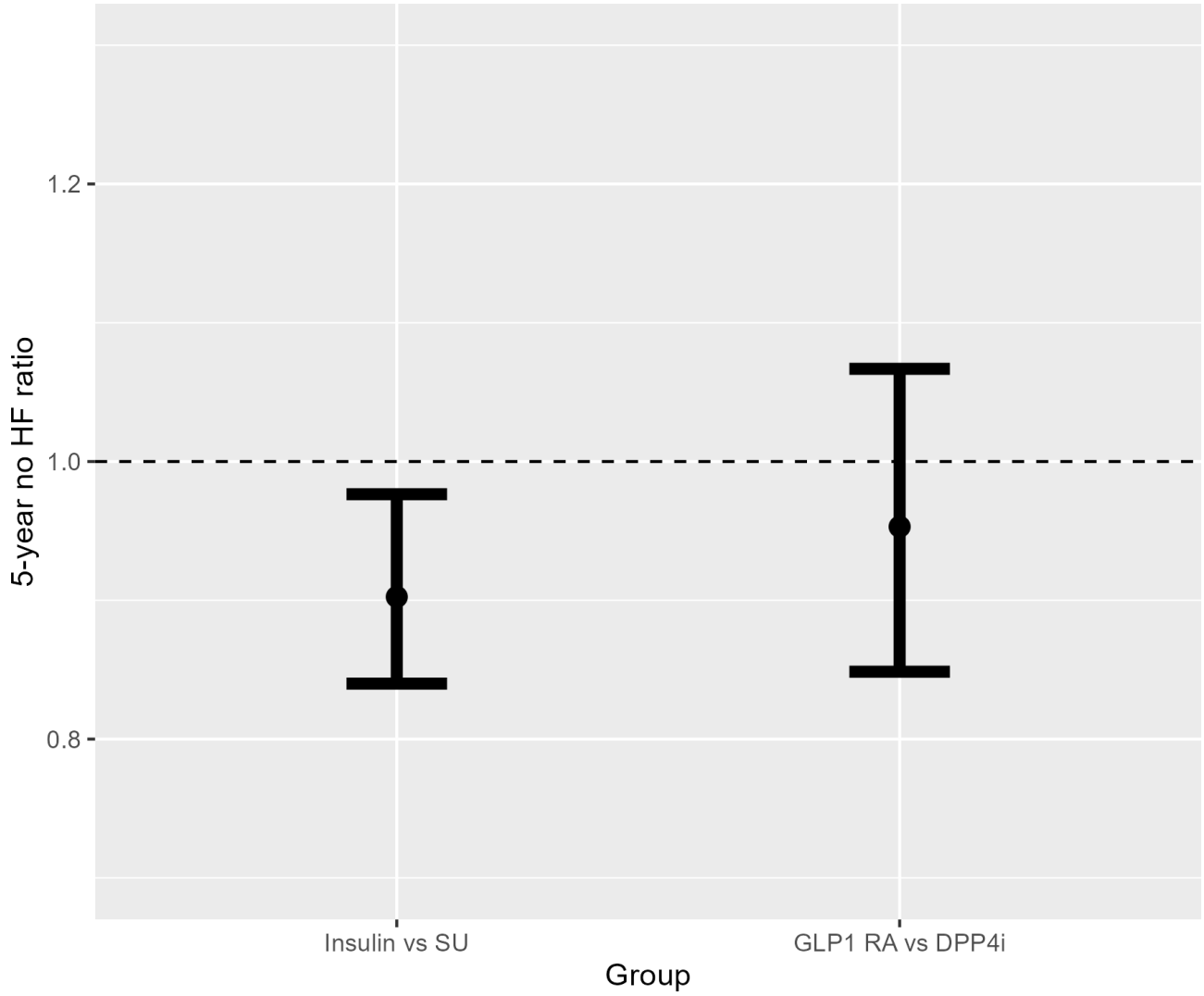


Figure 4. Analysis results on HF-free survival ratios. DPP4i: dipeptidyl peptidase-4 inhibitors; GLP1 RA: glucagon-like peptide-1 receptor agonists; HF: heart failure; SU: sulfonylureas.



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

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.

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]

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]

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]

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

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

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

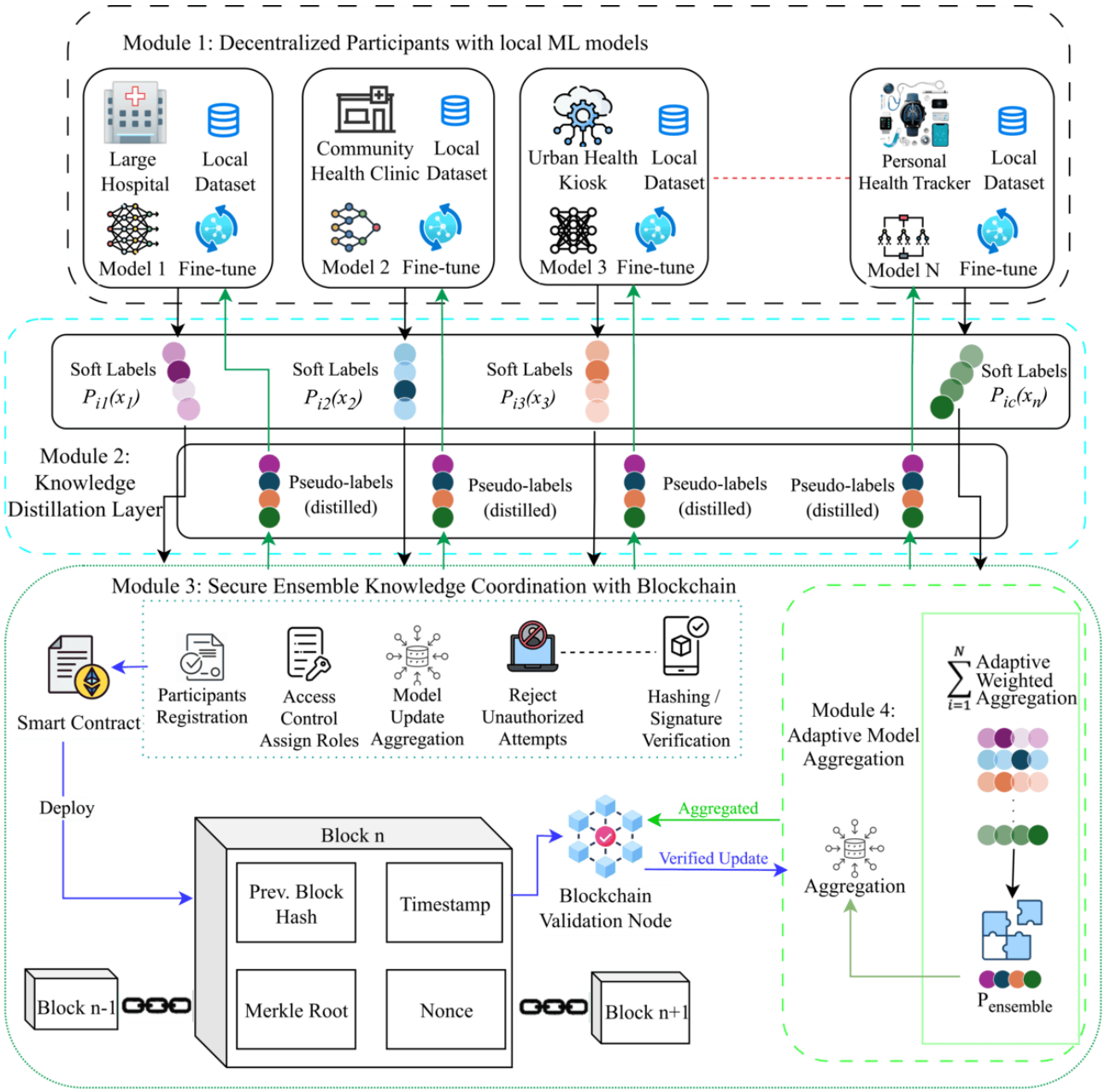
Overview of FedEnTrust

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.

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.

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

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:

$$(1)\quad Pi(x)=[p1,p2,\ldots,pc]\in Rc, \text{ where } \sum c=1Cpc=1$$

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:

$$(2)\quad Pi(t)(x)=\text{softmax}\ (zi(x)T),\ T=2$$

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:

$$(3)\quad Wi(t)=Acci(t)\sum j=1NAccj(t)$$

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.

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=0.1max_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).

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

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.

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.

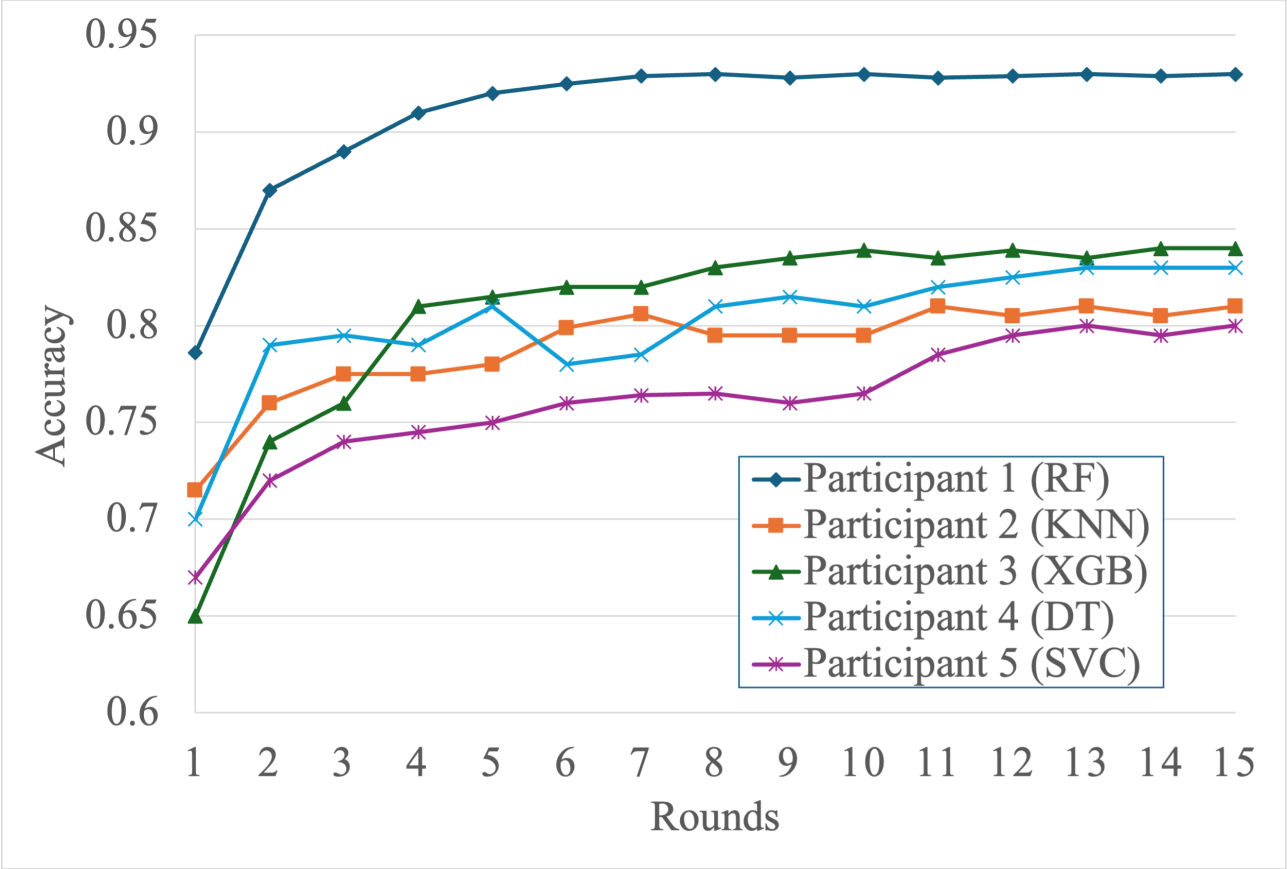


Figure 3 presents the corresponding model loss curves. All participants experienced substantial loss reduction early on, with convergence observed by round 15. Participant 1 maintained the lowest loss throughout, while participants 4 and 5 showed marked improvement from higher initial losses, demonstrating the benefit of federated collaboration.

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.

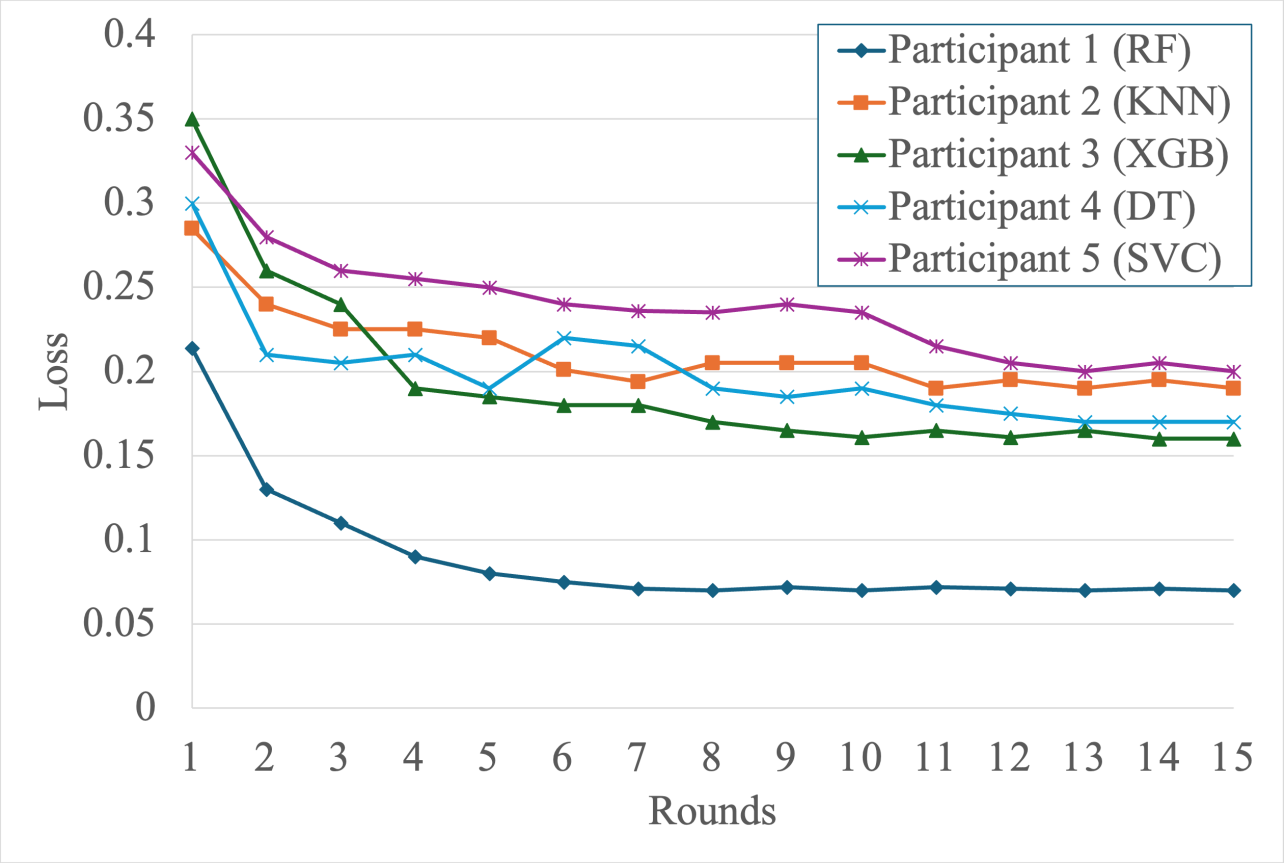


Table 5. Federated Models Performance after 15 rounds

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

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:

$$\Delta(b) = M_{\text{FedEnTrust}}(b) - M_{\text{Baseline}}(b) \quad (6)$$

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

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.

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.

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.

[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

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.

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]

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

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

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

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).

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.

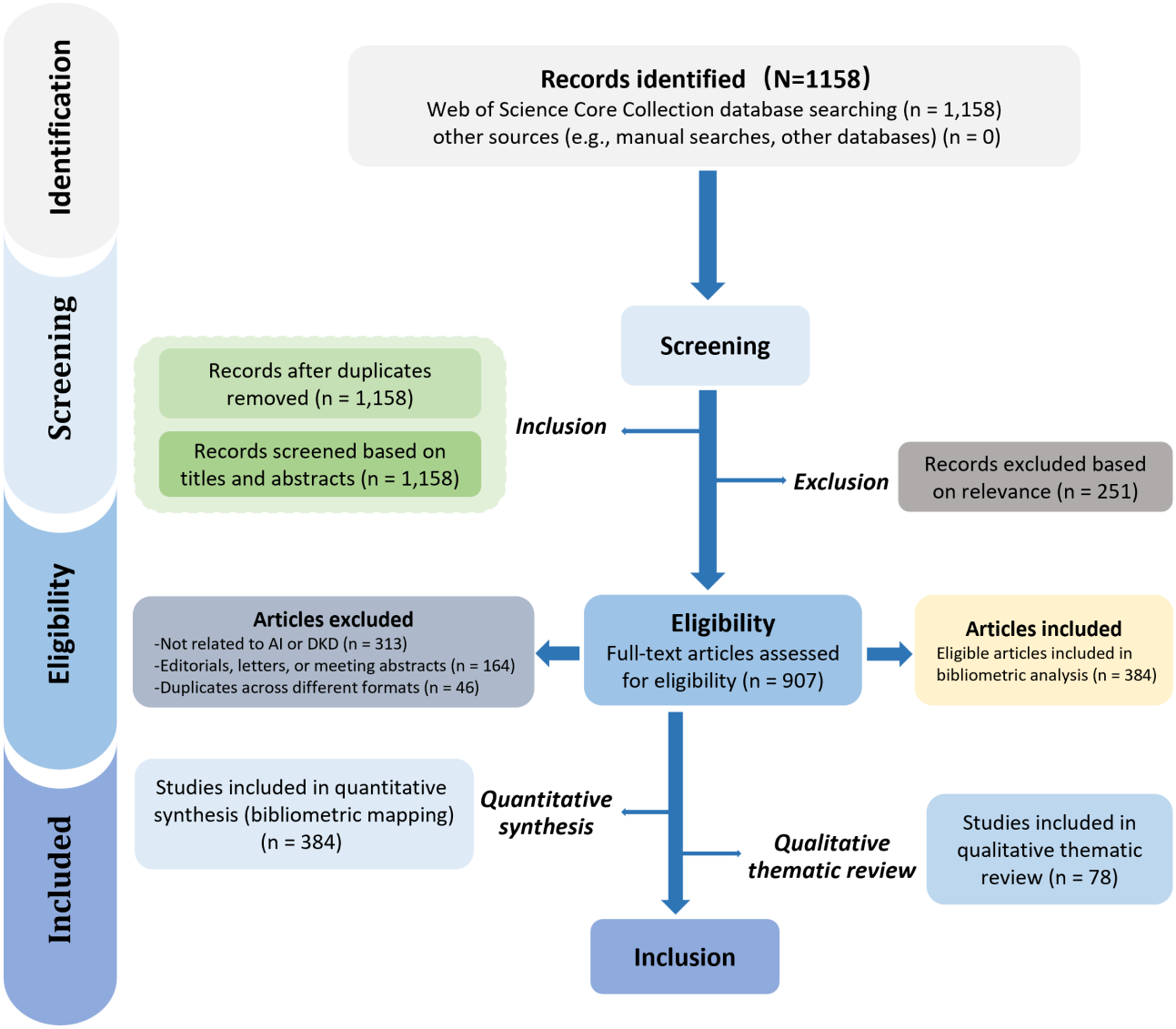
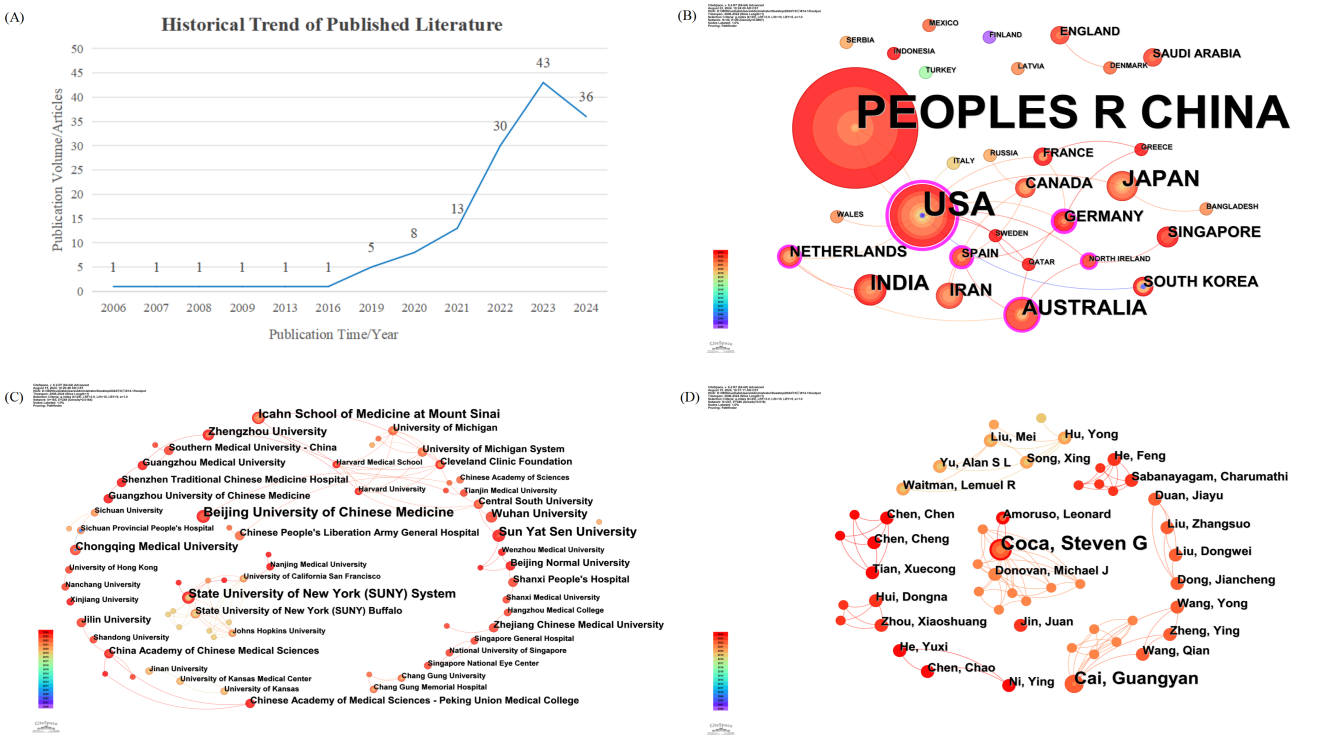


Figure 2. Analysis of the publication trends in artificial intelligence research on diabetic kidney disease from 2006 to 2024: (A) timeline of annual publications, (B) co-occurrence network of research countries, (C) co-occurrence network of research institutions, and (D) co-occurrence network of authors.



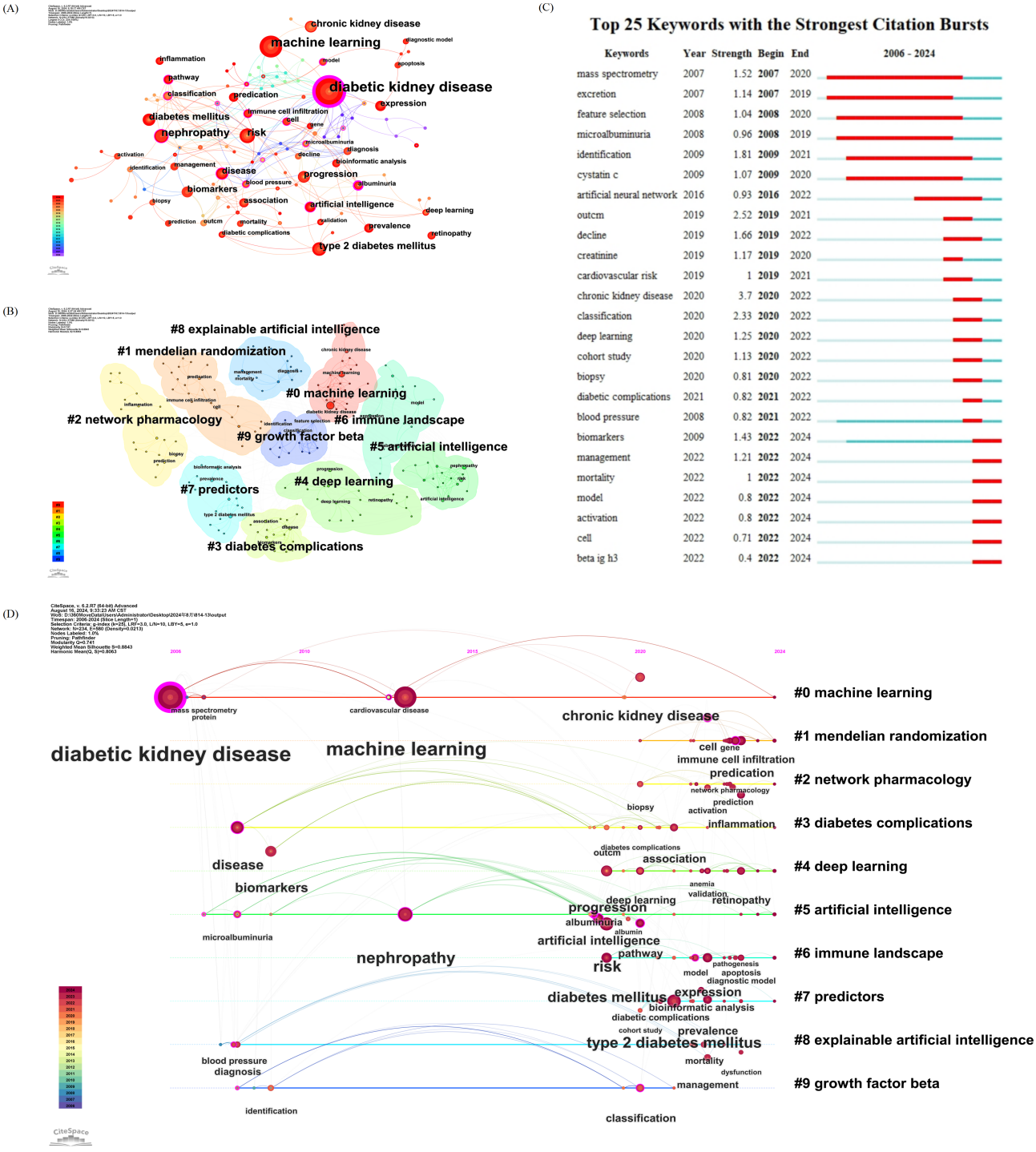
Geographic and Institutional Contributions

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).

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

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

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

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

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org