Contents

Original Papers

Analysis of Diabetes Apps to Assess Privacy-Related Permissions: Systematic Search of Apps (e16146)
José Flors-Sidro, Mowafa Househ, Alaa Abd-Alrazaq, Josep Vidal-Alaball, Luis Fernandez-Luque, Carlos Sanchez-Bocanegra. ........................................ 3

Role of Digital Engagement in Diabetes Care Beyond Measurement: Retrospective Cohort Study (e24030)
Yifat Fundoiano-Hershcovitz, Abigail Hirsch, Sharon Dar, Eitan Feniger, Pavel Goldstein. .................................................................................. 16

Using Virtual Reality to Improve Health Care Providers’ Cultural Self-Efficacy and Diabetes Attitudes: Pilot Questionnaire Study (e23708)
Elizabeth Beverly, Carrie Love, Matthew Love, Eric Williams, John Bowditch. ............................................................................................. 27

Feasibility of the Web-Based Intervention Designed to Educate and Improve Adherence Through Learning to Use Continuous Glucose Monitor (IDEAL CGM) Training and Follow-Up Support Intervention: Randomized Controlled Pilot Study (e15410)
Madison Smith, Anastasia Albanese-O’Neill, Yingwei Yao, Diana Wilkie, Michael Haller, Gail Keenan. ........................................................................... 38

Evaluation of a Diabetes Remote Monitoring Program Facilitated by Connected Glucose Meters for Patients With Poorly Controlled Type 2 Diabetes: Randomized Crossover Trial (e25574)
Daniel Amante, David Harlan, Stephenie Lemon, David McManus, Oladapo Olaitan, Sherry Pagoto, Ben Gerber, Michael Thompson. .................. 52

Exchanges in a Virtual Environment for Diabetes Self-Management Education and Support: Social Network Analysis (e21611)
Carlos Pérez-Aldana, Allison Lewinski, Constance Johnson, Allison Vorderstrasse, Sahiti Myreni. ........................................................................ 76

Early Insights From a Digitally Enhanced Diabetes Self-Management Education and Support Program: Single-Arm Nonrandomized Trial (e25295)
Folasade Wilson-Anumudu, Ryan Quan, Cynthia Castro Sweet, Christian Cerrada, Jessie Juusola, Michael Turken, Carolyn Bradner Jasik. ........ 8

Diabetes Engagement and Activation Platform for Implementation and Effectiveness of Automated Virtual Type 2 Diabetes Self-Management Education: Randomized Controlled Trial (e26621)
Roy Sabo, Jo Robins, Stacy Lutz, Paulette Kashiri, Teresa Day, Benjamin Webel, Alex Krist. ................................................................. 96

Ability of Current Machine Learning Algorithms to Predict and Detect Hypoglycemia in Patients With Diabetes Mellitus: Meta-analysis (e22458)
Satoru Kodama, Kazuya Fujihara, Haruka Shiozaki, Chika Horioka, Mayuko Yamada, Takaaki Sato, Yuta Yaguchi, Masahiko Yamamoto, Masaru Kitazawa, Midori Iwanaga, Yasuhiro Matsubayashi, Hirohito Sone. ........................................ 108
Public Perspectives on Anti-Diabetic Drugs: Exploratory Analysis of Twitter Posts (e24681)
Su Golder, Millie Bach, Karen O'Connor, Robert Gross, Sean Hennessy, Graciela Gonzalez Hernandez. 136

Diabetes Distress and Glycemic Control in Type 2 Diabetes: Mediator and Moderator Analysis of a Peer Support Intervention (e21400)
Kara Mizokami-Stout, Hwajung Choi, Caroline Richardson, Gretchen Platt, Michele Heisler. 165

Using Wearable Activity Trackers to Predict Type 2 Diabetes: Machine Learning–Based Cross-sectional Study of the UK Biobank Accelerometer Cohort (e23364)
Benjamin Lam, Michael Catt, Sophie Cassidy, Jaume Bacardit, Philip Darke, Sam Butterfield, Ossama Alshabrawy, Michael Trenell, Paolo Missier. 177

Reviews

Telemetric Interventions Offer New Opportunities for Managing Type 1 Diabetes Mellitus: Systematic Meta-review (e20270)
Claudia Eberle, Stefanie Stichling. 65

Experiences of Young People and Their Caregivers of Using Technology to Manage Type 1 Diabetes Mellitus: Systematic Literature Review and Narrative Synthesis (e20973)
Nicola Brew-Sam, Madhur Chhabra, Anne Parkinson, Kristal Hannan, Ellen Brown, Lachlan Pedley, Karen Brown, Kristine Wright, Elizabeth Pedley, Christopher Nolan, Christine Phillips, Hanna Suominen, Antonio Tricoli, Jane Desborough. 122

Application of the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies in Assessing Mobile-Delivered Technologies for the Self-Management of Type 2 Diabetes Mellitus: Scoping Review (e23687)
Jessica Forsyth, Hannah Chase, Nia Roberts, Laura Armitage, Andrew Farmer. 149
Analysis of Diabetes Apps to Assess Privacy-Related Permissions: Systematic Search of Apps

José Javier Flors-Sidro¹, MSc; Mowafa Househ², PhD; Alaa Abd-Alrazaq², PhD; Josep Vidal-Alaball³⁴, MD, MPH, PhD; Luis Fernandez-Luque⁵⁶, PhD; Carlos Luis Sanchez-Bocanegra⁷, PhD

¹Information Systems Department, Consorci Hospitalari Provincial de Castelló, Castelló de la Plana, Spain
²Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
³Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain
⁴Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la recerca a l’Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain
⁵Salumedia Labs, Seville, Spain
⁶Adhera Health Inc, Palo Alto, CA, United States
⁷Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Alaa Abd-Alrazaq, PhD
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University
Education City
Doha,
Qatar
Phone: 974 55708549
Email: aabdralrazaq@hbku.edu.qa

Abstract

Background: Mobile health has become a major vehicle of support for people living with diabetes. Accordingly, the availability of mobile apps for diabetes has been steadily increasing. Most of the previous reviews of diabetes apps have focused on the apps’ features and their alignment with clinical guidelines. However, there is a lack of knowledge on the actual compliance of diabetes apps with privacy and data security guidelines.

Objective: The aim of this study was to assess the levels of privacy of mobile apps for diabetes to contribute to the raising of awareness of privacy issues for app users, developers, and governmental data protection regulators.

Methods: We developed a semiautomatic app search module capable of retrieving Android apps’ privacy-related information, particularly the dangerous permissions required by apps, with the aim of analyzing privacy aspects related to diabetes apps. Following the research selection criteria, the original 882 apps were narrowed down to 497 apps that were included in the analysis.

Results: Approximately 60% of the analyzed diabetes apps requested potentially dangerous permissions, which pose a significant risk to users’ data privacy. In addition, 28.4% (141/497) of the apps did not provide a website for their privacy policy. Moreover, it was found that 40.0% (199/497) of the apps contained advertising, and some apps that claimed not to contain advertisements actually did. Ninety-five percent of the apps were free, and those belonging to the “medical” and “health and fitness” categories were the most popular. However, app users do not always realize that the free apps’ business model is largely based on advertising and, consequently, on sharing or selling their private data, either directly or indirectly, to unknown third parties.

Conclusions: The aforementioned findings confirm the necessity of educating patients and health care providers and raising their awareness regarding the privacy aspects of diabetes apps. Therefore, this research recommends properly and comprehensively training users, ensuring that governments and regulatory bodies enforce strict data protection laws, devising much tougher security policies and protocols in Android and in the Google Play Store, and implicating and supervising all stakeholders in the apps’ development process.

(JMIR Diabetes 2021;6(1):e16146) doi:10.2196/16146)
KEYWORDS
diabetes mellitus; privacy; mobile apps; dangerous permissions

Introduction

Background

Diabetes mellitus (DM) is one of the most common chronic conditions around the globe. The number of people with DM has risen globally from 108 million in 1980 to 422 million in 2014 [1]. Its prevalence has been increasing everywhere, especially in middle-income countries, from 4.7% in 1980 to 8.5% in 2014. DM increases the risk of serious health problems such as myocardial infarction, renal failure, stroke, and lower limb amputation [2]. Diabetic retinopathy is one of the most important causes of blindness worldwide, especially in developed countries [3]. DM has also been linked to an increased risk of other conditions such as dementia, depression, and some types of cancer [4]. In order to reduce the risk of complications, intensive patient education and support are needed, which can be enhanced by the use of mobile technology.

Along with the exponential increase in the number of health apps [5,6], in particular the number of diabetes apps has increased significantly in the last several years [7]. Mobile health (mHealth) has become a major vehicle of support for people living with diabetes, and the availability of mobile apps for diabetes has been steadily increasing. Most of the previous reviews of diabetes apps have focused on their features and their alignment with clinical guidelines [8,9]. However, there is a lack of knowledge on the actual compliance of diabetes apps with privacy and data security guidelines.

Therefore, there is a growing concern to review diabetes apps because in many cases they do not possess the quality and content that they should according to their own declared purposes [10,11]. In addition, some studies that have investigated the effectiveness of mobile apps clearly demonstrate data privacy problems [12], as well as a lack of transparency with the provided information [13]. Studies on mHealth and privacy have raised some serious concerns in recent years. Because very sensitive information is increasingly accessed and shared using mobile apps, there is an obvious need for clinicians, software developers, users, and patients to be aware of and trained on information privacy aspects. Personal data may be collected through different means, such as being entered directly by the user or being recorded by the phone’s camera, microphone, or paired wireless device (e.g., Bluetooth glucometer apps). It is crucial to note that the treatment of these critical data demands a special approach regarding security and privacy. However, some apps do not even provide information regarding their privacy policies. In some instances, these privacy terms are difficult to understand by nontechnical users, and some privacy policies may even be regarded as abusive. To make matters worse, the ecosystem of mobile apps is so complex that even app developers and users may not know with whom the data is being shared and for what purpose [14-16].

The aim of this study was to evaluate the privacy-related permissions of Android diabetes apps in Google’s Play Store using a semiautomatic approach that relies on the extraction of privacy-related features (e.g., permissions, terms of usage). This approach was designed to assist in identifying strategies to raise the awareness of app users, patients, and clinicians. To illustrate our approach, we provide two case studies of diabetes apps that were comprehensively analyzed (Multimedia Appendix 1).

Methods

Study Design

The first step in this study was the extraction of metadata from mobile apps’ metadata using a web-based application programming interface (API) [23]. We used the platform 42Matters, which offers a web-based commercial tool that facilitates access to the Android Google Play Store and to other mobile platforms’ apps’ metadata through a proprietary API [24]. Searches were conducted with the developed script module 42Matters’ index of Android apps. Since the 42Matters platform did not allow the extraction of privacy-related permissions from Apple’s App Store, the research centered on Android apps from Google’s Play Store. Data extraction was focused on potentially dangerous permissions [25] that allow the requesting app access to private user data or control over the mobile device, both of which can negatively impact the user. Because this type of permission introduces potential risk, the system does not automatically grant it to the requesting app. Our methodology was based on similar studies of health apps that used the 42Matters platform, but focusing on privacy-related information [26,27].

In order to complement the quantitative results already presented, we described and investigated two very popular and well-rated diabetes apps (presented in Multimedia Appendix 1) from a qualitative perspective.

http://diabetes.jmir.org/2021/1/e16146/
For the extraction of the diabetes apps’ metadata, we first devised the architecture [28] and subsequently developed the corresponding software module for the automatic extraction of mobile app metadata using the web-based API of 42Matters. The output of this module is a data set stored locally in a comma-separated values (CSV) file. The source code for the module was released under the GNU AGPLv3 license and can be found on the GitHub link [29]. This module is capable of querying the API of the 42Matters platform to retrieve metadata related to diabetes apps, including the Android permissions required by the apps. The module was designed to extract apps with the following search parameters: (1) language (we searched for English-language apps), (2) keyword search (we searched for apps whose titles included the root words “diabet” and “mellitus”), and (3) app categories (we selected the categories medical, health and fitness, lifestyle, and education).

The resulting apps were manually reviewed (see Multimedia Appendix 1) to assess whether they were related to diabetes. All apps were related to diabetes, but we did not address the quality of their content. As explained in the “Limitations” section, choosing a method where search fields matched the description—and not only the title—would have resulted in more apps, many of which would not have been related to diabetes.

Once the most suitable app categories were identified, it was then possible to move on to design the entire app selection process, which consisted of the following steps (see Figure 1):
Figure 1. App selection process flowchart.

- Step 1: “Identification” phase—all of the diabetes apps that contained the root words “diabet” or “mellitus” in an app’s title field were selected, resulting in 882 apps; by matching diabet or mellitus, it was possible to ensure that any relevant potential variations of the words that contained these root words (ie, diabetes, diabetic, diabetics, mellitus, etc) were included in the search.
- Step 2: “Category filtering” phase—in order to guarantee that only relevant diabetes apps were included in the study, all the retrieved apps that did not belong to the medical, health and fitness, education, or lifestyle categories [30] were automatically filtered out by the 42Matters script module and excluded from the study; this filtering resulted in 732 apps.
- Step 3: “Screening” phase—in this phase, we manually filtered apps and excluded 5 diabetes apps related to pets, 1 discontinued app, and 55 duplicated apps; this screening resulted in 671 apps.
- Step 4: “Eligibility” phase—we excluded apps that did not have a minimum of 50 downloads, and therefore discarded 174 apps.
• Step 5: “Inclusion” phase—the resulting 497 apps were analyzed, which were the objects of analysis of this research.

Data Extraction: Retrieved Metadata Fields

After the final set of apps was selected in June 2019, a process was initiated to extract all the relevant metadata and information, which were stored in a CSV file. All the retrieved fields are described in the table below.

Table 1. Description of apps’ retrieved metadata as provided by 42Matters.

<table>
<thead>
<tr>
<th>App’s metadata field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Main name of the app</td>
</tr>
<tr>
<td>Price</td>
<td>Price and currency (0 if it was free)</td>
</tr>
<tr>
<td>Permission</td>
<td>Required Android permissions of the app</td>
</tr>
<tr>
<td>Rating</td>
<td>App’s average rating from 0 to 5 (0=worst, 5=best)</td>
</tr>
<tr>
<td>Number of downloads</td>
<td>Number of times the app was downloaded</td>
</tr>
<tr>
<td>Number of ratings</td>
<td>Number of times the app was rated</td>
</tr>
<tr>
<td>Contains advertising</td>
<td>True if the app contained advertising and false if it did not</td>
</tr>
<tr>
<td>Category</td>
<td>Category to which the app belonged (medical, health and fitness, education, or lifestyle)</td>
</tr>
<tr>
<td>Short description</td>
<td>Short description of the app’s declared purpose</td>
</tr>
<tr>
<td>Website</td>
<td>Website of the app</td>
</tr>
<tr>
<td>Privacy policy</td>
<td>Website showing the app’s privacy policy</td>
</tr>
</tbody>
</table>

Extraction of Android Privacy-Related Permissions

Starting with Android 6.0 (API 23 level), users grant permissions to apps while using them, not when an app is installed. On the one hand, this approach simplifies the process of installing the app because the user does not need to grant permissions when installing or updating the app. In addition, it provides the user with more control over the app’s functionalities because users can revoke the granted permissions from the app’s configuration screen at any time. On the other hand, this new approach complicates the app’s usability because dangerous permissions have to be granted while using the app, which poses an additional challenge for untrained users. Android distinguishes between 4 categories of permissions: normal, signature, dangerous, and special [31].

Signature and special permissions will not be explained here because they are rarely used and were not found in any of the apps included in our research. The most frequently requested permissions are normal and dangerous permissions. If an app declares a normal permission in its manifest, the system grants permission to it automatically without the user’s intervention. On the other hand, Android considers dangerous permissions as critical because they allow apps to access users’ critical data. More concretely, an Android dangerous permission [25,32] allows the requesting app access to private user data or control over the mobile device. Because this type of permission allows developers to access users’ data, photos, and videos stored on the device, it introduces potential risk, and the system does not automatically grant it to the requesting app [33,34].

In brief, normal permissions do not put the user’s privacy at risk directly. Consequently, if an app declares a normal permission in its metadata, the system grants permission to it automatically without the user’s intervention. On the other hand, a dangerous permission allows an app to access the user’s critical data, and consequently the user should explicitly authorize this permission [35]. The 10 most required dangerous permissions found in this research are shown in Multimedia Appendix 2.

Results

App Functions

The process described in the “Methods” section retrieved a total of 497 apps (Multimedia Appendix 3). The breakdown of privacy-related permissions is summarized in Table 2. Most of the apps required at least one dangerous permission.

Table 2. Summary of the privacy-related main features of retrieved diabetes apps.

<table>
<thead>
<tr>
<th>Assessed parameter</th>
<th>Diabetes apps (N=497), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not require any permissions (either normal or dangerous)</td>
<td>89 (17.9)</td>
</tr>
<tr>
<td>Only requires normal permissions</td>
<td>111 (22.3)</td>
</tr>
<tr>
<td>Requires at least one dangerous permission</td>
<td>297 (59.8)</td>
</tr>
<tr>
<td>Does not provide a website link to its privacy policy</td>
<td>141 (28.4)</td>
</tr>
<tr>
<td>Contains advertising</td>
<td>199 (40.0)</td>
</tr>
</tbody>
</table>
The reason for apps not requesting any permissions is that they serve very basic functions (e.g., calculators, logs, diaries, etc) that only need access to very basic and noncritical Android resources. Only 22.3% (111/497) of the apps required normal (noncritical) permissions alone. On the other hand, 59.8% (297/497) of the apps required at least one dangerous permission. This might be partially justified by these apps’ more advanced functionalities (e.g., doctor-patient interaction, connecting to a glucometer, calorie-burning calculation, scanning the barcode of diabetic food, etc).

Regarding privacy, it was worrying to discover that 28.4% (141/497) of the apps did not return the privacy policy metadata field, consequently posing additional difficulty for users to adequately understand how these apps would treat very sensitive personal information.

Finally, 40.0% (199/497) of the apps contained advertising, which can imply the sharing of critical personal data (e.g., a user’s precise location) with unknown third parties for geolocated advertisement. Consequently, because the advertising business model in the mobile ecosystem is usually linked to the sharing or selling of critical personal data [36], the aforementioned findings unquestionably confirm the necessity to educate users and raise awareness regarding user privacy in diabetes apps.

### Dangerous Permissions

As explained below, dangerous permissions refer to permissions that might lead to data breaches of private information [37]. From the 497 diabetes apps included in our final analysis, a substantial number of them—297 (59.8%)—required dangerous permissions. Table 3 shows, in decreasing order, which dangerous permissions were most frequently requested by the apps.

<table>
<thead>
<tr>
<th>Dangerous permission</th>
<th>Diabetes apps that requested it (N=497), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write external storage</td>
<td>272 (54.7)</td>
</tr>
<tr>
<td>Read external storage</td>
<td>169 (34.0)</td>
</tr>
<tr>
<td>Access coarse location</td>
<td>103 (20.7)</td>
</tr>
<tr>
<td>Access fine location</td>
<td>95 (19.1)</td>
</tr>
<tr>
<td>Camera</td>
<td>89 (17.9)</td>
</tr>
<tr>
<td>Get accounts</td>
<td>82 (16.5)</td>
</tr>
<tr>
<td>Read phone state</td>
<td>81 (16.3)</td>
</tr>
<tr>
<td>Record audio</td>
<td>39 (7.8)</td>
</tr>
<tr>
<td>Call phone</td>
<td>23 (4.6)</td>
</tr>
<tr>
<td>Read contacts</td>
<td>22 (4.4)</td>
</tr>
<tr>
<td>Others (the sum of the remaining dangerous permissions)</td>
<td>28 (5.6)</td>
</tr>
</tbody>
</table>

In addition, Figure 2 illustrates the number of apps that required each of the top 14 dangerous permissions, arranged by category. The four quadrants represent each of the four categories to which the apps belonged: education, health and fitness, medical, and lifestyle. In addition, the “Advertising” tag indicates whether an app contained advertising: the ones in blue contained advertising, while the ones in red did not. The x-axis shows the number of apps, while the y-axis lists the 14 most requested dangerous permissions.
Figure 2. The top 14 dangerous permissions by app category (lifestyle, medical, education, and health and fitness) and type of privacy-related permission requested, as well as whether they included advertising (“True”) or not (“False”).

Discussion

Principal Results and Comparison With Previous Work

Although we identified the apps requesting access to the camera (89/497, 17.9%), we need to study the actual usage of apps in order to fully understand the context before we consider that access to be a potential risk. For instance, in the case of diabetes, it is very common to use the camera for food logging. On the other hand, except for advertising or fitness tracking (e.g., calorie counting), the need for the user’s geolocation data seems difficult to justify. In this sense, what might be acceptable in one app might not be reasonable in others. Similar studies found that 77 of 186 (41.4%) permissions requested by 58 popular German mHealth apps were not related in any way to the apps’ functionalities [38]. Moreover, 15 of 42 (35.7%) Android health and well-being apps accredited by the UK’s NHS Health Apps Library requested critical permissions for unjustifiable reasons [12]. Similarly, other research concluded that several popular mental health apps and mHealth apps requested permissions that were not aligned with the apps’ stated purposes [14,21]. One of the consequences of requesting unnecessary dangerous permissions is a decrease in users’ trust, acceptance, and use of these apps.

Another finding of this study was that 95.4% of the apps were free of charge. The business model of free apps is, in most cases, based on advertising (through services such as Google AdMob), resulting in the disclosure of users’ critical data, either directly (through the app itself) or indirectly (through Google’s commercial advertising platforms).

The reliance on advertising of some of the studied apps might be linked to the high number of apps requesting geolocation, since location can increase advertisement revenue. A study on
NHS-accredited apps found some evidence that patients’ data were information for advertisers [12]. Other studies also found that users’ information was shared in 19 of 24 popular medication-related apps in the United Kingdom, the United States, Canada, and Australia [39]. Research of privacy in the top 36 mental health and smoking cessation apps also found a lack of compliance with disclosing or sending data to third-party providers [40]. Although app developers usually claim that they do not collect or share personally identifiable data, users can be easily identified by correlating advertising services using data analytics [39]. In addition, 28.4% of the studied apps did not provide a privacy policy website, which corroborates results from other research that demonstrated that 48% of 17,991 free Android apps did not have a privacy policy [18]. Building on this finding, 81% of 154 Android apps related to hypertension and diabetes did not refer to a privacy policy [33]. In addition, a privacy policy was missing in 417 of 600 (69.5%) prominent mHealth apps [41]. Most likely, had we not discarded less reliable apps in our research, the percentage of apps that did not provide a link to a website with their privacy policy would have been higher [34]. The lack of a privacy policy is a critical fault, as it prevents users from properly understanding how apps treat their very sensitive personal information. Further, the discrepancy between apps’ privacy policies and their actual features has been reported in several studies [12,18]. This issue might be partially attributed to the fact that app developers have insufficient knowledge about privacy best practices [42]. In our study, 59.8% of apps required at least one dangerous permission, the two most requested being write external storage (54.7%) and read external storage (34.0%). This finding confirms the results from previous research. For instance, the most common dangerous permissions requested by the most popular freeware mHealth apps were write external storage (90%) and read external storage (50%) [34]. For prominent mental health apps in the Google Play Store, the most frequently requested permissions were also write (73%) and read (73%) external storage. In addition, these two permissions were the most requested (79%) in medicine-related apps in the Google Play Store in the United Kingdom, the United States, Canada, and Australia [38]. These permissions may indeed jeopardize users’ privacy because they allow developers to access users’ data, photos, and videos stored on the device [33,34]. Another relevant finding was that health and fitness apps usually requested more dangerous permissions than apps belonging to other categories [21]. Apps’ ever-changing functionality and privacy policies, as well as their complexity, do not facilitate matters, either. Moreover, having to manually accept dangerous permissions when using an app poses an additional challenge that can have detrimental consequences, particularly for less knowledgeable users. For instance, individuals with low literacy rates or the elderly would require adequate training to truly understand what they are consenting to before using diabetes apps. Existing tools to evaluate eHealth literacy skills [43] do include security awareness as one of their dimensions. However, the complexity of potential security issues is increasing, and it might be necessary to develop new tools and training methods for both patients and health care providers.

**Practical Implications**

These findings have very important practical implications for users, physicians, developers, and policy makers [44,45]. To select an appropriate mobile app for diabetes, end users should be aware of what type of personal data is collected, used, and shared by a certain app by carefully reading the app’s description, terms of use, and privacy policy. In addition, it is imperative to emphasize the need for training so that users are able to understand complex privacy policies and terms of service and are fully aware of the privacy risks derived from the sharing of their data with third parties. Users should also be knowledgeable about the different types of dangerous permissions so that they can discern how each particular permission may jeopardize their data. The ultimate goal is to empower users so that they can autonomously and proficiently deny access to any unjustifiable dangerous permission.

To minimize the privacy risks derived from using diabetes apps, savvy users should use AdBlock or encryption apps [33]. Moreover, health care providers should ensure that the apps they recommend to patients adhere to a strict privacy code, and they should assist users in selecting suitable apps by explaining both the apps’ benefits and their risks.

App developers should enforce their apps’ full compliance with internationally recommended standards and practices [46-49]. Specifically, developers must ensure that their apps’ privacy policies are always readily available, very simple to read, and able to be understood by any user. Further, their apps should never request dangerous permissions not directly related to the apps’ declared purpose. Developers should not—without the users’ explicit consent—collect, use, or share user data for any purpose outside of the predefined scope of the app, and all data sharing practices should be transparently disclosed to users. Last but not least, developers should be aware of diverse privacy laws and data protection legislation, which differ greatly depending on the country or region of use.

In terms of privacy laws, apps tend to adhere to the data protection legislation in the developers’ country of origin but not in the apps’ country of use. Therefore, regulators around the world should collaborate to establish a specific international accreditation program for diabetes apps. Such a program should be based on unified privacy best practices in which user privacy is the main priority. Because app developers reserve the right to change their privacy policies at any given time and modify their apps’ declared purpose and functionalities, regulators should regularly monitor developers’ adherence to the recommended privacy practices. As well, regulators should emphasize developers’ responsibility and accountability for protecting user data. In addition, app stores should mandate stringent principles and standards that actually compel developers to provide simple and intelligible privacy policies in their apps, especially taking into consideration untrained or illiterate users.
Limitations

We opted to use the free version of the commercial platform 42Matters instead of the Google Play Store because the Google Play Store had a limit of 250 apps per query.

Another limitation was that the developed module exclusively searched for all diabetes apps that contained the root words diabet or mellitus in the title field. There are some diabetes apps in which the aforementioned root words appear in the app’s description but not in the app’s name. Therefore, some diabetes-related apps may have been excluded from the study. However, this criterion was selected for two principal reasons: (1) to ensure that only truly diabetes-related apps were retrieved, and (2) to make the best use of limited resources (there was neither enough time nor enough labor to thoroughly screen 4700+ apps, many of which bore no relation whatsoever to diabetes). In this sense, our research was not intended to be exhaustive. Rather, we wanted to quantify and evaluate the overall privacy characteristics of the most representative sample of diabetes-related apps. A broader search (ie, to query for all apps that contained the root words diabet or mellitus in the apps’ descriptions) would certainly have yielded many false positives of apps unrelated to diabetes and hence required a very resource-intensive manual screening of the apps, which would have been an unnecessary complication of the overall analysis process.

The study did not comprehensively address either the fact that the number of permissions an app requests does not necessarily reflect how risky the app may be. For instance, an app requesting, unnecessarily, a single dangerous permission, could seriously endanger users’ personal data by collecting and illegitimately sharing them. On the other hand, an app requesting multiple dangerous permissions, but for valid technical or functional needs, could be considered safe. Therefore, the amount of personal information that users are putting at risk depends on many factors, such as the app’s functionality, the permissions it requests, and the context in which these permissions are being used [50]. To perform a more complete assessment of apps’ privacy risks, additional technical, human, and contextual research (eg, analysis of the skills of patients using diabetes apps) should be conducted. For example, when dealing with privacy issues in health apps, an important factor to be considered would be the legitimacy of the request, as highlighted in a recent publication on mHealth apps for cancer in which the authors evaluated a new scale to assess the privacy policies of mHealth apps [51]. Tracking users’ location might be fair in the case of reporting a medical emergency (eg, hypoglycemic crisis).

Although the methodology employed in this research was robust and Google is continuously improving Android and the Play Store’s security policy, this study found evidence that it is extremely difficult to prove whether diabetes apps actually comply with their privacy policies. In fact, even Google cannot control the many malicious apps that are frequently uploaded by hackers in its Play Store and is consequently forced to periodically remove massive numbers of these fraudulent apps [52-54]. Further, a recently published two-year study discovered 2040 potential counterfeit apps that contained malware in the Google Play Store [55].

This study did not cover all of the elements related to the privacy and security of diabetes apps. Privacy protection cannot be guaranteed solely by controlling permissions; for instance, unsecure internet connections can also jeopardize the privacy of mobile app users. Finally, our study only evaluated the apps on one app store; the privacy policies and the requested dangerous permissions in other app stores, such as Apple’s App Store or Samsung’s Galaxy Store, might have yielded different outcomes. However, Android’s Google Play Store was also chosen due to its popularity.

Future Research

A possible expansion of the research could include investigating those diabetes apps that were excluded from this research, either because they belonged to nonrelevant categories or because the developed module did not search for the root words in the apps’ description field. Future research could also focus on analyzing the taxonomy of app categories and match them to officially recognized and standardized clinical categories, such as the Systematized Nomenclature of Medicine Clinical Terms or Medical Subject Headings. Related to that, there is a new trend emerging toward the creation of machine learning approaches to identify privacy issues in mobile apps [56,57]. However, to the best of our knowledge, those methods have unfortunately not yet been applied to health apps. Further, there is a need for homogenous approaches for the assessment of privacy in health apps, as was highlighted recently in a scoping review addressing the issue [58].

Finally, from a legal perspective, although many diabetes apps are available worldwide, their privacy policies usually only comply with the specific national data protection regulations of the developers’ country or region of origin. For instance, the BeatO SMART Diabetes Management app claims that both its privacy policy and its terms of use fully adhere to Indian law, but if this app were to be used in the Middle East or the European Union, it would be unclear whether it would also comply with data protection laws in the country or region of use. This could indeed be another matter of study.

Conclusions

If privacy issues in diabetes mobile apps are not dealt with carefully, users may unwillingly and unknowingly share very sensitive private data. Therefore, it is crucial that all stakeholders are involved in the development of diabetes apps from the very beginning of the process in order to ensure apps’ absolute compliance with data protection laws and user privacy. As the economic value of personal data increases [59], a completely new business model for apps has emerged: users pay for the usage of an app with their data, which is then sold to third parties, such as advertising clients [60]. The lesson to be learned is that there is a price to pay in exchange for free apps, usually at the expense of privacy. Consequently, new control measures are needed to enable users to decide which personal information they are willing to disclose in return for a certain service [61].
The importance of personal data protection laws and their endorsement are of utmost importance. Well-designed privacy policies may protect individuals by requiring consent for the collection, use, disclosure, or retention of sensitive personal and health information, and they may regulate the use of these extremely sensitive data, allowing users to modify their information as well as to revoke their previous consent. Therefore, we recommend proper training for users, enforcement of strict data protection laws by governments and regulatory bodies, much tougher security policies and protocols in both Android apps and the Google Play Store, and the implication and supervision of all stakeholders in the app development process.

Authors' Contributions

JF-S was the principal investigator. He designed the majority of the work, supervised the research, and took over most of the data interpretation and writing of the manuscript. In addition, he was responsible for developing the software module for extracting apps’ metadata. MH and AA-A significantly contributed to the results and discussion sections of the paper. JV-A contributed to the overall manuscript and study by providing a clinical perspective. LF-L conceived the original research idea and greatly assisted with the design of the methodology and with the discussion section. Finally, CLS-B’s contribution to the analysis and interpretation of the results was fundamental. All of the authors contributed to and approved the manuscript.

Conflicts of Interest

LF-L is co-founder of Adhera Health Inc (USA), a digital health company that provides digital therapeutic solutions for people with chronic conditions.

Multimedia Appendix 1
Qualitative results of case studies.
[DOCX File, 5315 KB - diabetes_v6i1e16146_app1.docx]

Multimedia Appendix 2
Top 10 Android’s dangerous permissions identified.
[DOCX File, 16 KB - diabetes_v6i1e16146_app2.docx]

Multimedia Appendix 3
Comma-separated values files.
[DOCX File, 14 KB - diabetes_v6i1e16146_app3.docx]

References


Abbreviations

API: application programming interface
CSV: comma-separated values
DM: diabetes mellitus
mHealth: mobile health
Original Paper

Role of Digital Engagement in Diabetes Care Beyond Measurement: Retrospective Cohort Study

Yifat Fundoiano-Hershcovitz, PhD; Abigail Hirsch, PhD; Sharon Dar, MSc; Eitan Feniger, BSc; Pavel Goldstein, PhD

DarioHealth, Caesarea, Israel

School of Public Health, University of Haifa, Haifa, Israel

Corresponding Author:
Yifat Fundoiano-Hershcovitz, PhD
DarioHealth
Hatochen, 8
Caesarea, 3088900
Israel
Phone: 972 525296979
Email: Yifat@mydario.com

Abstract

Background: The use of remote data capture for monitoring blood glucose and supporting digital apps is becoming the norm in diabetes care. One common goal of such apps is to increase user awareness and engagement with their day-to-day health-related behaviors (digital engagement) in order to improve diabetes outcomes. However, we lack a deep understanding of the complicated association between digital engagement and diabetes outcomes.

Objective: This study investigated the association between digital engagement (operationalized as tagging of behaviors alongside glucose measurements) and the monthly average blood glucose level in persons with type 2 diabetes during the first year of managing their diabetes with a digital chronic disease management platform. We hypothesize that during the first 6 months, blood glucose levels will drop faster and further in patients with increased digital engagement and that difference in outcomes will persist for the remainder of the year. Finally, we hypothesize that disaggregated between- and within-person variabilities in digital engagement will predict individual-level changes in blood glucose levels.

Methods: This retrospective real-world analysis followed 998 people with type 2 diabetes who regularly tracked their blood glucose levels with the Dario digital therapeutics platform for chronic diseases. Subjects included “nontaggers” (users who rarely or never used app features to notice and track mealtime, food, exercise, mood, and location, n=585) and “taggers” (users who used these features, n=413) representing increased digital engagement. Within- and between-person variabilities in tagging behavior were disaggregated to reveal the association between tagging behavior and blood glucose levels. The associations between an individual’s tagging behavior in a given month and the monthly average blood glucose level in the following month were analyzed for quasicausal effects. A generalized mixed piecewise statistical framework was applied throughout.

Results: Analysis revealed significant improvement in the monthly average blood glucose level during the first 6 months ($t=-10.01, P<.001$), which was maintained during the following 6 months ($t=-1.54, P=.12$). Moreover, taggers demonstrated a significantly steeper improvement in the initial period relative to nontaggers ($t=2.15, P=.03$). Additional findings included a within-user quasicausal nonlinear link between tagging behavior and glucose control improvement with a 1-month lag. More specifically, increased tagging behavior in any given month resulted in a 43% improvement in glucose levels in the next month up to a person-specific average in tagging intensity ($t=-11.02, P<.001$). Above that within-person mean level of digital engagement, glucose levels remained stable but did not show additional improvement with increased tagging ($t=0.82, P=.41$). When assessed alongside within-person effects, between-person changes in tagging behavior were not associated with changes in monthly average glucose levels ($t=1.30, P=.20$).

Conclusions: This study sheds light on the source of the association between user engagement with a diabetes tracking app and the clinical condition, highlighting the importance of within-person changes versus between-person differences. Our findings underscore the need for and provide a basis for a personalized approach to digital health.

(JMIR Diabetes 2021;6(1):e24030) doi:10.2196/24030

http://diabetes.jmir.org/2021/1/e24030/
KEYWORDS
blood glucose; mHealth; diabetes; self-management; digital engagement

Introduction

Diabetes mellitus is characterized by hyperglycemia that can reduce life expectancy [1], cause considerable health complications, increase cost of care, and lower quality of life [2, 3]. The treatment of diabetes mellitus is challenging for both persons with diabetes and clinicians because successful management requires sustained patient-driven lifestyle changes [4, 5]. For many, the fundamental challenge of managing chronic diabetes is doing what is needed rather than knowing what to do per se. Research suggests that patients need more than theoretical knowledge about healthy eating, exercise, and self-monitoring of blood glucose [6]. They also need assistance building awareness of their daily health-related behaviors. This awareness building and engagement with prohealth behaviors seeds the implementation of a prohealth lifestyle [7-10].

Technology-driven solutions can help persons with type 2 diabetes bridge the gap between knowing what to do, building awareness and engagement, and implementing these changes [11, 12]. Mobile apps have been shown to improve diabetic outcomes via education and support for adhering to evidence-based recommendations [13-16]. Apps for diabetes management and diabetes online communities appear to be useful tools for helping people with type 2 diabetes to control HbA1c and are increasingly considered core intervention tools in self-management for patients with type 2 diabetes [17-19].

Such apps often include the following two core features: a method for recording blood glucose measurements and a vehicle for logging behaviors and situations that impact health outcomes. Paper-and-pencil logging of activities, such as meals, food intake, and exercise, alongside blood glucose measurements has been a long-standing best practice for building awareness and helping individuals better control their glucose levels. In the emerging world of digital diabetes care, tagging (creating a digital in-app activity log) represents a convenient alternative for activity tracking that can be leveraged for app-based diabetes self-management [20].

Health behavior change theory posits that new health behaviors emerge when people gain both knowledge and self-efficacy to implement the said knowledge [21-23]. We posit that the moment of marking (tagging) one’s context in conjunction with taking a blood glucose measurement is a prime opportunity for reinforcing knowledge and building self-efficacy. It is possible that what is being tagged is of less importance than the act of tagging something. In other words, by tagging with measurement, persons with type 2 diabetes transform each glucose reading into a moment of quick reflection on their context and actions proceeding that measurement. This moment of focused awareness building may be a key piece in launching a virtuous process of improved future health behavior.

However, as the usage of apps to capture blood glucose data and to log behavior increases, sophisticated analysis of the rich data now available has lagged. Research gaps include understanding the general blood glucose trajectory among persons with type 2 diabetes using digital diabetes support tool users, the association between app engagement and short- and long-term clinical outcomes, and the relative impact of specific app features dedicated to self-management [11, 15, 24]. In addition, strikingly little work has focused on disentangling the value of remote digital capture of glucose measurements versus digital engagement via tagging. Nuanced modeling of the impact of different features within diabetes apps could help to maximize the impact of mobile health apps on behavior change and, by extension, on health outcomes [25]. Of note, previous studies suggested that changes in diabetes clinical outcomes appear to have the following two phases: an initial improvement over 6 months, followed by a longer-term sustained period [26, 27]. Modeling that allows for a multitrajectory process, that is, for change trajectories to have different slopes at different periods of time, while not the norm in many assessments of digital health platforms, seems imperative.

Over the last decade, behavioral science research has increasingly focused on between-person processes as opposed to within-person processes [28]. Surprisingly, the quantitative literature on diabetes still generally emphasizes treatment efficacy and associated between-person group-level factors and ignores within-person variability [29-31]. However, disaggregating between-person and within-person variability can illuminate the dynamics of the relative contribution of intraperson changes versus between-person differences to successful diabetes management. Moreover, this kind of analysis enables testing quasicausal relationships by adding lagged effects between modeled within-person digital engagement and clinical outcomes. Finally, as described above, the associations between digital engagement and clinical outcomes are not necessarily linear, as has been mostly assumed previously [32].

This study leverages a retrospective analysis of a home-use diabetes glucometer with full data capture in a supporting mobile app among type 2 diabetes patients with poorly controlled blood glucose levels. We hypothesized that during the first 6 months of using a chronic condition self-management app, tagging alongside blood glucose measuring would be associated with reduced blood glucose levels. By modeling the two-stage trajectory process, we expected to show the improvement to persist until the end of the 1-year study period. We also hypothesized that disaggregated within- and between-person variabilities in engagement behaviors would be predictive of reductions in monthly average blood glucose levels. Moreover, we suspected that 1-month lagged within-person digital engagement would be associated with improvements in monthly average blood glucose levels.

Methods

Platform

This study utilized the Dario digital therapeutics solution for chronic diseases to support self-management of diabetes. The Dario platform combines an innovative meter with a phone app...
that is available for both Android and iOS devices. The glucometer consists of a small pocket-sized holder for strips, a lancet, and the meter. The meter is removed from the holder and plugged directly into a cell phone, effectively converting the cell phone into the display screen for the meter. Connecting the meter directly to the phone has two advantages. First, it ensures 100% data capture during glucose readings. Second, it means users have opened the mobile app with each glucose measurement. This makes contextually tagging a measurement very easy to do at the time of taking the measurement. More specifically, the glucose meter is physically attached to the mobile phone, and the measurement is shown on the mobile phone (the meter does not have a screen) in a “decision support system” view. After the measurement is shown, a data entry screen is presented, where additional information can be added. The additional information includes measurement time (fasting/premeal/postmeal/bedtime); carbohydrate intake (grams); meal, mood, and location settings; and physical activity (kcal). All information is stored in the patient log book in the app “attached” to the specific blood glucose reading. Data are uploaded to the cloud for backup and further analysis, as presented in Figure 1. An extended version of this figure is provided in Multimedia Appendix 1.

**Figure 1.** Dario mobile app platform. (A) Data entry screen allows tagging measurement type, carbohydrate intake (grams), physical activity (kcal), and tags such as mood setting and location. (B) Logbook screen presenting measurements and tagging records.

**Measures**

The monthly average blood glucose level, which was defined as the mean of all of a user’s blood glucose measurements taken over a 30-day interval, was used as the core outcome metric. Independent variables included digital engagement, operationalized as the number of times a user added a tag to a measurement each month, and available demographic variables of gender and age. All data were transferred and stored in compliance with Health Insurance Portability and Accountability Act (HIPAA) requirements, using Amazon AWS database services. All data were anonymized before extraction for this study.
Users

The 998 users included in this analysis used the Dario platform between 2016 and 2020. The inclusion criteria were as follows: type 2 diabetes, noninsulin treatment, first month blood glucose average >180 mg/dL, blood glucose measurements during the first 2 months on the system, and at least five blood glucose measurements during the first and 12th months on the platform.

Users were grouped by their use of the behavioral tagging features of the app. The “taggers” group included users with an average of more than one tag per month over the 12-month activity (n=413). Users who only used the app for blood glucose measurements were designated as “nontaggers” with an average of one or less than one tag per month over the 12-month activity (n=585).

No difference between the groups was found for gender ($\chi^2=0.19, P=.66$), age ($B=0.96, t_{596}=1.20, P=.23$), initial blood glucose level ($B=5.89, t_{596}=1.64, P=.10$), and the average number of monthly blood glucose measurements over the study period ($B=-0.26, t_{595}=-0.18, P=.85$).

Ethical & Independent Review Services [33], a professional review board, issued the institutional review board exemption for this study (18032-03#).

Analytical Approach

Statistical analysis was conducted in two stages. The first stage modeled differences in the monthly average blood glucose level throughout users’ initial 12 months on the Dario platform, grouped by taggers and nontaggers. The second analysis focused on the association between disaggregated within- and between-patient tagging behaviors and the monthly average blood glucose level. The test was two-tailed.

First Analysis: Testing Differences in the Monthly Average Blood Glucose Level Throughout the Initial 12 Months by Taggers and Nontaggers

The standard linear longitudinal model assumes a single slope growth pattern for changes in an outcome variable across time. Sometimes, such a simple model does not fit the empirical data. In contrast, piecewise - based mixed - effects models allow flexibility in the modeling of variable change trajectories across time [34]. Here, a mixed piecewise model assessed differences in the monthly average blood glucose level in two segments (1-6 months and 7-12 months) with users grouped as taggers and nontaggers. The piecewise model allowed the data to exhibit different linear trends over their different regions. This statistical approach provided an opportunity to model curvilinear changes in the monthly average blood glucose level as a single process and to test complex effects based on this more flexible model. Based on previous research [26], the piecewise cutoff point for the model slopes was chosen at 6 months, assuming a change in the time-related monthly average blood glucose trajectory after 6 months of Dario device usage. We tested several residual distributions of the model outcome (Gaussian, log normal, and gamma) and different combinations of random effects. The model with the best fit, and thus used in the analysis, was based on log - normal residuals, and it included person-based random intercepts and random slopes for both periods (1-6 months and 7-12 months). The model also included an interaction between the groups (taggers and nontaggers) at both periods.

Second Analysis: Assessing Within-Person and Between-Person Associations Between Tagging Behavior and the Monthly Average Blood Glucose Level

The second analysis was performed on the entire sample of users (n=998), with a focus on continuous behavioral tagging within individuals as opposed to trends over time by groups in the first analysis. The monthly overall tagging volume was disaggregated to separate within- and between-person variabilities using person-level centering and person-level aggregation [29]. In addition, 1-month lagged tagging engagement was calculated based on the within-person engagement. Thereafter, a generalized mixed model assuming log-normal outcome residual distribution was applied to test the association of monthly within-person engagement and between-person engagement with the monthly average blood glucose level. The model also included 1-month lagged within-person engagement to test for a quasicausal relationship between a user’s tagging engagement and the monthly average blood glucose level. Since lagged engagement demonstrated a nonlinear relationship with the monthly average blood glucose level, a quadratic term for lagged engagement was also added to the model.

Finally, we tested a curve-linear pattern of the association between lagged within-person engagement and the monthly average blood glucose level by applying a piecewise generalized mixed model defining two slopes for the relationship with a cutoff point in the person-level mean of the lagged engagement.

Results

First Analysis: Piecewise Generalized Mixed Model Analysis

Patients’ age ($B=0.001, t=0.87, P=.38$) and gender ($B=-0.02, t=-1.61, P=.11$) were not related to the monthly average blood glucose level.

Piecewise mixed model analysis revealed a significant monthly average blood glucose decrease for both taggers ($B=-0.027, 95\% CI \text{-}0.033 \text{ to } -0.022$; monthly average blood glucose decrease=13%) and nontaggers ($B=-0.020, 95\% CI \text{-}0.024 \text{ to } -0.015$; monthly average blood glucose decrease=9%) during the period of the first 6 months of use (Figure 2). In addition, the monthly average blood glucose level showed significantly better improvement among taggers than among nontaggers ($B=0.008, 95\% CI 0.001 \text{ to } 0.014; t=2.15, P=.03$). Extended information is provided in Multimedia Appendix 2. During the period from 7 to 12 months, there were no significant time-related trending monthly average blood glucose levels among taggers ($B=-0.005, 95\% CI \text{-}0.014 \text{ to } 0.001$; monthly average blood glucose decrease=3%) and nontaggers ($B=-0.004, 95\% CI -0.011 \text{ to } 0.002$; monthly average blood glucose decrease=2%). Taggers and nontaggers likewise did not show significant differences in their time-related monthly average blood glucose trend ($B=0.001, 95\% CI \text{-}0.008 \text{ to } 0.011; t=0.29, P=.77$) during the second time period (7-12 months).
Second Analysis: Within- and Between-Person Associations Between Tagging and Health Conditions

The second analysis focused on the relationship between tagging behaviors and blood glucose levels, decoupling between- and within-person effects as opposed to trends over time examined in the first model. Within-person change in tagging activity was negatively associated with the monthly average blood glucose level ($B=-0.002$, $95\%$ CI $-0.0023$ to $-0.016$; $t=-2.15$, $P=0.03$) (Figure 3). Extended information is provided in Multimedia Appendix 3. Moreover, preceding month tagging showed a quadratic relationship with the monthly average blood glucose level. Finally, aggregated (between-subject) digital engagement was not related to the monthly average blood glucose level ($B=0.0005$, $95\%$ CI $-0.0003$ to $0.0012$; $t=1.30$, $P=0.20$).
Figure 3. Association between within-person 1-month lagged digital engagement and monthly average blood glucose (BG) (mg/dL). The blue line shows locally weighted smoothing with a 95% confidence interval (the surrounding dark grey area). The dotted gray line indicates results from the generalized mixed piecewise model with two slopes (below and above the person-level mean).

For a better understanding of the nonlinear effect that was found between preceding month digital engagement and the absolute monthly average blood glucose level, a piecewise generalized mixed framework was adopted for modeling two slopes of the relationship (below the person-level engagement mean and above the mean) (Figure 3). Up to the subject-level mean, preceding month digital engagement showed a negative association with the monthly average blood glucose level, resulting in a 43% monthly average blood glucose decrease ($B$=−0.004, 95% CI −0.005 to −0.003; $t$=−11.02, $P$<.001). Above the subject-level mean, preceding month digital engagement was not related to the monthly average blood glucose level ($B$=0.0002, 95% CI −0.0003 to 0.0008; $t$=0.82, $P$=.41).

To better understand the contribution of the single component of digital engagement to the association with blood glucose, we reran the model described above and included measurement time tagging (fasting/premeal/postmeal/bedtime); carbohydrate intake tagging (grams); meal, mood, and location settings; and physical activity tagging (kcal) instead of aggregated tagging. Based on the model, up to the subject-level mean, preceding month carbohydrate intake; meal time tagging; and meal, mood, and location settings showed negative associations with the monthly average blood glucose level ($B$=−0.004, $t$=−3.47, $P$<.001; $B$=−0.007, $t$=−5.56, $P$<.001; and $B$=−0.004, $t$=−6.29, $P$<.001, respectively). Above the subject-level mean, preceding month carbohydrate intake; meal time tagging; and meal, mood, and location settings were not related to the monthly average blood glucose level ($B$=0.002, $t$=1.53, $P$=.13; $B$=−0.0001, $t$=−0.14, $P$=.89; and $B$=−0.0001, $t$=−0.14, $P$=.89, respectively).

Physical activity tagging showed a similar result pattern but did not reach statistical significance (up to the subject-level mean: $B$=−0.001, $t$=−1.07, $P$=.28; above the subject-level mean: $B$=0.004, $t$=0.08, $P$=.93).

Discussion

Principal Results

This real-world analysis presents data analyzing associations between blood glucose levels and digital engagement (tagging) in a digital app for chronic health condition management. More specifically, the results indicate that two distinct phases exist for remote blood glucose monitoring via an app (a rapid improvement phase lasting about 6 months and then a maintenance phase, which was here followed to 12 months).
Moreover, the improvement is stronger for users with increased tagging behavior. In addition, disaggregating within- and between-person variabilities in digital engagement, we demonstrated the quasicausal relationship between within-person behavioral tagging in any given month and the blood glucose level in the following month.

Consistent with the literature, we found that users of a connected glucose monitor experienced the most change in their first few months of use [14,15,27]. Of note, change patterns with an early rapid change period followed by a long-tailed period where change is retained appeared in many real-world digital interventions for behavior change [35,36]. While findings of a pre-post intervention change that remains stable after intervention are expected in traditional structured time-bound interventions, most digital health interventions are continuous in nature and thus might be believed to follow a smoother trajectory [37]. Nonetheless, evidence is emerging that there is a distinctly different impact in the short term versus the longer term, even for continuous eHealth interventions. This study shows that utilization of a piecewise mixed model statistical framework appears to be the more appropriate base model to describe a user’s two-phase slope change in blood glucose levels. Likewise, utilization of a piecewise approach allows independent analysis of predictors and covariates for the adoption versus longer-term periods. The piecewise-based model indicates that during the short-term adoption phase, while both taggers and nontaggers show declines in average blood glucose levels, taggers show significantly steeper declines than nontaggers. In other words, tagging appears to build behavioral awareness to life management, contributing to the glucose balance [38]. However, in the longer term, at 7 to 12 months, both groups evidenced flat trajectories, suggesting that over the long term, gains are sustained and durable but not increasing. Building behavioral awareness by means of a digital therapeutics platform addresses barriers to diabetes self-care in the context of everyday life. Previous studies revealed that behavior engagement is associated with increased individual diabetes-related problem-solving ability and with significant improvement in glucose control. Similar to our findings, these improvements were sustained at long-term follow-ups [37,39]. Indeed, following 12 months, the improved glucose level in the taggers group persisted and remained lower than that in the nontaggers group.

Another distinct feature of digital therapeutics is the potential to deliver highly person-centric care. Personalized medicine has been called the “new mantra” in health care [40]. Here too, a move beyond the standard between-subject statistical approach is called for. Disaggregating within- and between-person variabilities in digital engagement enabled evaluation of the association between digital engagement and the monthly average blood glucose level, and in fact, only the within-person component had a significant contribution in predicting the blood glucose level in this model.

Moreover, we demonstrated the quasicausal relationship between within-person behavioral tagging in any given month and the blood glucose level in the following month by applying a piecewise-based mixed model owing to the nonlinear nature of this association. We found a significant lagged association between digital engagement and the monthly average blood glucose level. Increased digital engagement was related to better clinical outcomes when digital engagement was below the person-level average (up to 43% improvement). However, above the person-level average, no association was observed. Here, between-person behavior engagement had no association with the monthly average blood glucose level. In other words, the within-subject component, as opposed to the between-subject component, is the source of the relationship between digital engagement and the blood glucose level.

Recent reviews call for research that moves beyond looking at “do digital health applications work” to more nuanced investigations that disentangle the relative contributions of active ingredients in digital health management protocols [13]. Our findings indicate that the strongest lever for helping people to lower their blood glucose levels is to ensure that they tag each month at least to the level of their personal critical tagging inflection point. Based on these findings, it turns out that just simple boosting of digital engagement to the maximum is not an efficient way to optimize glucose levels in diabetes patients. However, tracking digital engagement for persons with type 2 diabetes and maintaining it just around their average may result in optimal levels of glucose and reduction in patient efforts and digital fatigue. We expect that the analytical approach applied in this study will be beneficial for personalizing interventions and optimizing incentivization planning.

This information could be used to further personalize outreach and incentivization efforts to encourage users to maintain their personal critical level of tagging. At the same time, tagging above the personal mean yields no additional benefit in terms of current or future monthly average blood glucose levels. In other words, messaging that pushes for more tagging is unlikely to drive better glucose levels.

Limitations
We note several limitations in this study. First, as in all studies involving retrospective real-world data, groups were not randomly assigned and treatment protocols were not prescribed. Both factors create challenges for drawing causal effects. It certainly is possible that people who chose to tag behaviors were those who were the most motivated to change. Our inclusion criteria were designed to ensure that both taggers and nontaggers showed evidence of being motivated about their diabetes care. Fingerstick for regular blood glucose measurement certainly has a higher demand on time and energy than adding a few behavioral tags. All people included in this study were performing measurements regularly over the 12-month period of the study, and there were no differences between groups in terms of the volume of measurements. This would suggest that motivation may not be the primary difference between taggers and nontaggers. At the same time, this also limits the extendibility of the findings to low-measuring and thus presumably low-motivation populations. That said, the within-person analysis of lagged association covers the pitfalls of the classical between-group design, focusing on intrapersonal changes and allowing a quasicausal inference.

In this real-world data analysis, the time scale was designed to reflect monthly interval change over a 12-month period.
However, the relationships of interest in this study could be potentially investigated in different scales emphasizing daily, weekly, or monthly fluctuations. Owing to the difficulty in tracking daily changes in digital engagement in real-world studies, most studies focus on monthly fluctuations. Investigating fine-grained measurements with microintervals for tagging would certainly contribute to the literature [31]. Another challenge regarding our data was that available demographic data were limited. While there were no between-group differences by age or gender and no impacts of age and gender on the models, uncontrolled demographic biases might have been present from these or other demographic factors.

**Conclusions**

It appears highly likely that tagging features in a chronic condition management app, which are presented at the time of measurement, will help users with type 2 diabetes pause and pay attention to their daily life behaviors and connect these to their blood glucose measurements. Focusing on behavior and context as an integrated part of the glucose measurement process nearly doubled the clinical impact observed in users who only measured blood glucose. Likewise, while there was considerable variability in the volume of tagging, the more a user tagged in a given month, the lower the blood glucose level was likely to be in the next month until a user-specific threshold. Above that threshold, more tagging was not associated with a better clinical outcome.

From a behavioral science perspective, perhaps this is not so surprising. Directing focus onto actionable areas for improvement is likely to queue increased thought and action, and at the same time, the amount of attention to actionable areas needed is likely to vary considerably within individuals.

Future work investigating strategies beyond tagging that drive focus on and execution of actionable prohealth behaviors in a highly personalized within-person manner is certainly needed. Furthermore, similar studies examining piecemeal trajectories and within- versus between-person impacts of other behavior change tactics, including health coaching, gamification, and targeted tips, are warranted. Such a body of literature would help to move the field beyond the current state of “do digital tools work” to a nuanced understanding of what tools drive what clinical outcomes for which people under what circumstances.

**Conflicts of Interest**

YFH, AH, SD, and EF are employees of Dario Health. PG has received a consulting fee to assist with analyses but otherwise has no conflicts of interest.

**Multimedia Appendix 1**

Dario mobile app platform. Data entry screen allows tagging measurement time (fasting, premeal, postmeal, and bedtime); carbohydrate intake (grams); meal, mood, and location settings; and physical activity (kcal).

[PNG File, 201 KB - diabetes_v61e24030_app1.png]

**Multimedia Appendix 2**

Generalized piecewise mixed model for testing the differences in time-related monthly average blood glucose trajectories between taggers and nontaggers.

[DOCX File, 15 KB - diabetes_v61e24030_app2.docx]

**Multimedia Appendix 3**

Generalized piecewise mixed model for testing the association of within- and between-person engagement with the monthly average blood glucose level.

[DOCX File, 15 KB - diabetes_v61e24030_app3.docx]

**References**


http://diabetes.jmir.org/2021/1/e24030/
information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Using Virtual Reality to Improve Health Care Providers’ Cultural Self-Efficacy and Diabetes Attitudes: Pilot Questionnaire Study

Elizabeth Ann Beverly*1, PhD; Carrie Love2*, MFA; Matthew Love2*, MFA; Eric Williams2*, MFA; John Bowditch2*, MFA

1Department of Primary Care, Heritage College of Osteopathic Medicine, Ohio University, Athens, OH, United States
2Game Research and Immersive Design Lab, J Warren McClure School of Emerging Communication Technologies, Ohio University, Athens, OH, United States
*all authors contributed equally

Corresponding Author:
Elizabeth Ann Beverly, PhD
Department of Primary Care
Heritage College of Osteopathic Medicine
Ohio University
1 Ohio University
Grosvenor Hall 357
Athens, OH, 45701
United States
Phone: 1 17405934616
Fax: 1 7405932205
Email: beverle1@ohio.edu

Abstract

Background: In southeastern Appalachian Ohio, the prevalence of diabetes is 19.9%, nearly double that of the national average of 10.5%. Here, people with diabetes are more likely to have a delayed diagnosis, limited access to health care, and lower health literacy. Despite the high rates of diabetes in the region, the availability of endocrinologists and certified diabetes care and education specialists is limited. Therefore, innovative strategies to address the growing diabetes care demands are needed. One approach is to train the primary care workforce in new and emerging therapies for type 2 diabetes to meet the increasing demands and complexity of diabetes care.

Objective: The aim of this study was to assess the effectiveness of a virtual reality training program designed to improve cultural self-efficacy and diabetes attitudes.

Methods: Health care providers and administrators were recruited from large health care systems, private practices, university-owned hospitals or clinics, Federally Qualified Health Centers, local health departments, and AmeriCorps. Providers and administrators participated in a 3-hour virtual reality training program consisting of 360-degree videos produced in a professional, cinematic manner; this technique is called virtual reality cinema (cine-VR). Questionnaires measuring cultural self-efficacy, diabetes attitudes, and presence in cine-VR were administered to providers and administrators before and after the program.

Results: A total of 69 participants completed the study. The mean age of the sample was 42.2 years (SD 13.7), 86% (59/69) identified as female, 83% (57/69) identified as White, 86% (59/69) identified as providers, and 25% (17/69) identified as nurses. Following the training program, we observed positive improvements in all three of the cultural self-efficacy subscales: Cognitive (mean change –1.29; t65=–9.309; P<.001), Practical (mean change –1.85; t65=–9.319; P<.001), and Affective (mean change –0.75; t65=–7.067; P<.001). We observed the largest magnitude of change with the subscale, with a Cohen d of 1.16 indicating a very large effect. In addition, we observed positive improvements in all five of the diabetes attitude subscales: Need for special training (mean change –0.21; t67=–6.154; P<.001), Seriousness of type 2 diabetes (mean change –0.34; t67=–8.114; P<.001), Value of tight glucose control (mean change –0.13; t67=–3.029; P<.001), Psychosocial impact of diabetes (mean change –0.33; t67=–6.610; P<.001), and Attitude toward patient autonomy (mean change –0.17; t67=–3.889; P<.001). We observed the largest magnitude of change with the Psychosocial impact of diabetes subscale, with a Cohen d of 0.87 indicating a large effect. We observed only

http://diabetes.jmir.org/2021/1/e23708/
Introduction

Appalachia is a 205,000-square-mile region that encompasses 420 counties in 13 US states from Mississippi to New York. Ohio’s Appalachian region encompasses 32 counties [1], of which 16 are designated as economically at risk or distressed [2]. Here, 17.2% of the population live below the poverty line as compared to 14.4% for the rest of the state [3], and the counties with the highest poverty rates, ranging from 22.5% to 30.2%, are Appalachian [3]. People who live in Appalachian Ohio are more likely to be unemployed, have lower educational achievement, and limited access to transportation [4]. These social determinants of health contribute to the health disparities observed among people living in this region [5].

One health disparity disproportionately affecting people in Appalachian Ohio is diabetes [5]. An alarming 19.9% of adults in southeastern Ohio have diabetes [6], which is nearly double the national average of 10.5% [7]. In this region, people are more likely to have a delayed diabetes diagnosis, limited access to health care, lower health literacy, and lower empowerment [8,9]. For these reasons, people here are more likely to have macrovascular and microvascular complications, lower limb amputations, and depression [9-11]. Despite the high rates of diabetes in the region, the availability of endocrinologists and certified diabetes care and education specialists in Appalachian Ohio is limited [12]. Therefore, innovative strategies to address the growing diabetes care demands are needed.

One approach is to train the primary care workforce in new and emerging therapies for type 2 diabetes to meet the increasing demands and complexity of diabetes care. Primary care providers deliver more than 90% of the clinical care to people with type 2 diabetes in the United States [13]. This is even more pertinent in rural America where family physicians comprise a greater proportion of the workforce and provide comprehensive and irreplaceable care to the community [14]. Therefore, tailored continuing education for rural primary care providers and their staff is critical. Continuing education should address standards of medical care for diabetes as well as cultural competency and attitudes toward diabetes. Studies show that health care providers’ attitudes toward diabetes influence their approach to care (eg, paternalistic vs patient-centered care) and how they interact with people with diabetes [15-18]. Furthermore, continuing education that recognizes the unique cultural contributions of regions like Appalachian Ohio is necessary to improve providers’ ability to care for people from different backgrounds [19,20]. People from Appalachia share common language, behaviors, dietary habits, and value systems. Health care providers who understand their patients’ cultural backgrounds are more likely to observe improvements in diabetes outcomes and patient satisfaction [21,22]. Thus, tailoring continuing education to address diabetes attitudes and Appalachian culture is critical to improve the quality of care to an ever-increasing number of people with diabetes in Appalachian Ohio.

Virtual reality cinema (cine-VR) is an innovative educational technique that has the potential to transform the delivery and content of continuing medical education. Cine-VR is dynamic, accessible, and adaptable to providers’ needs and preferences [23]. Cine-VR gives providers access to life-like medical encounters without risk or harm to the patient. Further, cine-VR offers providers a glimpse into the lives of patients and culture of the region. These qualities are invaluable to geographically and culturally distinct regions like Appalachian Ohio.

For this study, we developed a 3-hour cine-VR training program designed to educate providers and administrators about diabetes, social determinants of health, and Appalachian culture. The aim of the study was to assess the effectiveness of cine-VR training in improving health care providers’ and administrators’ cultural sensitivity and diabetes attitudes. We hypothesized that cine-VR training would improve cultural self-efficacy and diabetes attitudes.

The following are our hypotheses:

1. Levels of cultural self-efficacy will increase after the 3-hour cine-VR training program.
2. Diabetes attitudes will improve after the 3-hour cine-VR training program.
3. Positive changes in cultural self-efficacy will be associated with increased presence in the cine-VR scenarios.
4. Positive changes in diabetes attitudes will be associated with increased presence in the cine-VR scenarios.

Methods

Overview

The purpose of this pilot study was to call attention to social determinants of health and Appalachian culture and to delineate their relationship to diabetes via 360-degree cine-VR simulations. Specifically, we administered questionnaires to providers and administrators before and after a cine-VR training program in order to (1) assess changes in cultural self-efficacy pre- and posttraining, (2) assess changes in diabetes attitudes pre- and posttraining, and (3) examine the relationship between changes in cultural self-efficacy and diabetes attitudes and...
presence in cine-VR. The Ohio University Office of Research Compliance approved the protocol (Institutional Review Board No. 19-X-99) and all recruitment procedures and materials.

**Recruitment**

Providers and administrators were recruited from large health care systems, private practices, university-owned hospitals or clinics, Federally Qualified Health Centers, local health departments, and AmeriCorps. In Appalachian Ohio, the majority of providers practiced at large health care systems and Federally Qualified Health Centers. Specifically, participants were recruited via emails from the Ohio University Diabetes Institute listserv and Area Health Education Center listserv, advertisements in social media, flyers in the community, and brief announcements at educational events. Participants included physicians, nurse practitioners, registered nurses, pharmacists, dietitians, certified diabetes educators, physical therapists, dentists, community health workers, and health care administrators and staff (eg, health department employees, free clinic directors, and AmeriCorps service members). The majority of providers specialized in primary care. Health care administrators were recruited given their role in health care–related decisions and their impact on quality of care. Additionally, administrators play a significant role in the assimilation of evidence-based management and training, and cine-VR has the potential to be an evidence-based educational training model.

**Power Analysis**

We conducted an a priori power analysis using Statuator [24], an online statistical calculator, which determined that a total sample size of 34 participants was estimated to achieve 80% power at a 5% significance level (P<.05) and to detect an effect size of 0.30.

**Cinematic 360-Degree Virtual Reality Simulations**

We hosted nine 3-hour training programs in Athens, Ohio. These training programs utilized 360-degree, virtual reality, professionally produced video in a cinematic manner to educate providers and administrators about diabetes, social determinants of health, and Appalachian culture. In the Using Virtual Reality to Visualize Diabetes in Appalachia program, participants watched 10 cine-VR simulations and two traditional films and observed interactions among the main character and her primary care physician, pharmacist, family, and community [25]. The main character in the simulations is Lula Mae, a 72-year-old woman with type 2 diabetes living in Appalachian Ohio. She is a widow; her husband died 27 years ago from a heart attack. She has three adult children and seven grandchildren. She cares full time for her adult son who suffered a traumatic brain injury from serving in the US Army. Lula Mae and her adult son live in an old house originally belonging to her grandparents. Her two adult daughters and grandchildren live on the same family land in their separate homes. Lula Mae is a source of care and support for her entire family, from her own children to her grandchildren. In doing so, her own health care needs come second to the daily needs of the people she loves. Despite Lula Mae’s struggles, we learn about the strengths of Appalachian culture and the resiliency one person can have if providers invest the time to connect with her one-on-one.

**Training Program Curriculum**

The Ohio University team developed a detailed curriculum taught synchronously with the cine-VR simulations. The curriculum included 12 modules that addressed the following content: (1) diabetes burnout, (2) food insecurity, (3) strengths of Appalachian culture, (4) rural transportation barriers, (5) elements of an effective patient–provider relationship, (6) diabetes and psychosocial issues, (7) high cost of diabetes medications, (8) gender roles in Appalachia, (9) cultural values in Appalachia, (10) diabetes complications, (11) diabetes comorbidities, and (12) patient-provider communication. An experienced behavioral diabetes researcher (EB) trained in interactive lecturing delivered all nine training sessions. The participants were encouraged to interact with each other and the lecturer. The lecturer incorporated straightforward and rhetorical questions to engage the participants. The simulations and curriculum were designed to increase cultural self-efficacy, improve diabetes attitudes, and increase presence in cine-VR. We provided 3.0 continuing medical education or continuing education credits for health care providers at no cost. Integrity of the education was ensured via a written curriculum, preapproved educational materials, and investigator observation of the training sessions.

**Virtual Reality Technology**

Working with the Ohio University’s Game Research and Immersive Design Lab, we leveraged a coalition of experts from Ohio University’s Diabetes Institute and the medical school, school of nursing, social work program, nutrition program, communication sciences and disorders program, school of film, theater program, and visual communication school. The interdisciplinary team consisted of one physician, three nurses, one social worker, one clinical psychologist, one audiologist, one registered dietitian, one health behaviorist, five filmmakers, four scriptwriters, and two website developers. This collaboration allowed us to create educational content that was not only medically accurate but emotionally powerful and visually stunning. Each series began with a traditionally shot short film to set the stage between Lula Mae and her relationship with a provider. This was followed by three cine-VR simulations that opened narrative windows into her daily life, her world, and her struggles. The fifth and sixth simulations of each series were guided simulations, a cine-VR face-to-face conversation with Lula Mae’s provider and Lula Mae herself. This six-video pattern was repeated twice, once covering Lula Mae’s relationship with her primary care provider and once covering her relationship with her local pharmacist.

The cine-VR simulations narratively demonstrated how Lula Mae’s social determinants of health and environment shaped her behaviors. Capturing those moments with camera systems that allow the audience to see a full 360-degree sphere created opportunities to present information in ways not possible with traditional filming methods. For example, when inside Lula Mae’s home, we saw the disorganization and chaos that resulted from a lack of social support. When the family car was stranded on the side of a remote road, we saw the transportation barriers
and isolation that families face in rural areas without public transportation. As a result of the 360-degree filming techniques employed, the team was able to present much more information about Lula Mae’s life and the factors affecting her diabetes.

The simulations were screened in an Oculus Go (Facebook Technologies) head-mounted display so that participants could turn their head and body in any direction and gather relevant information, much as if they were present in the actual location. Observant participants could notice subtle details, such as her surroundings, the condition of her home, or other activities co-occurring in the space. With traditionally shot films, this information would be presented in a close-up or with camera movement to call a viewer’s attention to relevant information, resulting in a more passive and guided viewing experience. Presenting the content in cine-VR creates an active viewing experience, with the viewer choosing what they want to watch and pay attention to, which increases immersion and encourages intellectual and emotional engagement. Viewers feel a sense of accomplishment as they notice subtle details planted by the filmmaking team, heightening the experience.

The fifth and sixth simulations of each series were what we called guided simulations, a prerecorded, cine-VR face-to-face conversation with Lula Mae’s provider and Lula Mae herself. Screened in a headset, these normally awkward, high-stakes conversations give the participants a chance to practice without the pressures of being watched or failing. Participants are encouraged to speak predetermined dialogue to a character in the headset and hear them respond. All of the cine-VR simulations were initiated simultaneously from a central computer, urging everyone in the room to say the same words at the same time, thereby reducing the potential for users to feel awkward about speaking aloud in public.

Measures
In addition to sociodemographic factors (ie, age, sex, race or ethnicity, occupation, years in practice, health care sector, percentage of Medicaid patients, and type of Medicaid patients), participants completed the following measures.

**Transcultural Self-Efficacy Tool–Multidisciplinary Healthcare Provider**

The Transcultural Self-Efficacy Tool–Multidisciplinary Healthcare Provider (TSET-MHP) is an 83-item scale that assesses changes in self-efficacy for cultural knowledge, cultural practical skills, and cultural awareness [26]. This scale yields three subscales: (1) Cognitive, (2) Practical, and (3) Affective [27]. All three subscales are rated on a 10-point scale, ranging from 1 (not confident) to 10 (totally confident). The Cognitive subscale asks participants to rate their level of confidence in their knowledge of the ways cultural factors influence health care for people belonging to different cultural backgrounds. The Practical subscale asks participants to rate their level of confidence in interviewing people of different cultural backgrounds to learn about their values, beliefs, and social determinants of health. Lastly, the Affective subscale asks participants to rate their level of confidence in acceptance of similarities and differences among cultural groups. These subscales demonstrate excellent internal consistency (Cronbach α ranging from .92 to .98) [27].

**Diabetes Attitude Scale-3**

The Diabetes Attitude Scale-3 (DAS-3) [17] is a 33-item scale that measures diabetes-related attitudes with five discrete subscales: (1) Need for special training (Cronbach α=.67), (2) Seriousness of type 2 diabetes (Cronbach α=.80), (3) Value of tight glucose control (Cronbach α=.72), (4) Psychosocial impact of diabetes (Cronbach α=.65), and (5) Attitude toward patient autonomy (Cronbach α=.76). Health care professionals are asked to rate their level of agreement on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The scale demonstrates good internal consistency and high content validity [17].

**Presence Questionnaire**

The 32-item Presence Questionnaire [28] measures the subjective experience of being in a virtual environment when a person is physically situated in another. Items are rated on a 7-point scale, ranging from 1 (not at all) to 7 (completely). We used a subset of 15 questions from the Witmer-Singer questionnaire and removed 17 questions that measured haptic (ie, the use of technology that simulates touch) factors because the cine-VR simulations did not involve interaction with the simulated environment. For example, we removed questions that asked participants about their ability to touch objects in the virtual environment or move around in the virtual environment (eg, “How closely were you able to examine objects?” or “How compelling was your sense of moving around inside the virtual environment?”) This revised questionnaire had four subscales: (1) Involvement (Cronbach α=.83), (2) Sensory Fidelity (Cronbach α=.75), (3) Adaptation and Immersion (Cronbach α=.46), and (4) Interface Quality (Cronbach α=.53). In addition, the research team added three questions to assess presence in the virtual environment; we labeled this fifth subscale Presence (Cronbach α=.78). We calculated our own internal consistency for each subscale using a reliability analysis. The revised 18-item questionnaire demonstrated internal consistency ranging from poor to very good.

**Data Collection**

At the training program, participants received a packet that included two copies of the informed consent form, a preassessment packet, and a postassessment packet. The principal investigator read the informed consent form to all attendees of the training program. Individuals interested in participating signed the informed consent form and placed it in the packet. The informed consent form emphasized the voluntary nature of participation and reminded participants that the study was not related to their participation in the overall training program. Participants completed a brief demographic form and the two preassessment questionnaires via pen and paper; this session lasted approximately 15 minutes. All questionnaires were prelabeled with an identification number prior to the start of the study. At the completion of the training program, participants completed three postassessment questionnaires via pen and paper; this session lasted approximately 15 minutes.
Participants with questions about the study were directed to email or call the principal investigator (EB).

**Statistical Analysis**

We assessed demographic factors using descriptive statistics and presented them as means and standard deviations or sample sizes and percentages. Chi-square tests, Fisher exact tests, independent-samples $t$ tests, and one-way analyses of variance were conducted to examine differences by age, gender, race, provider status, or percentage of Medicaid (ie, limited income and resources) patients. We performed paired $t$ tests to examine changes in TSET-MHP subscale scores and DAS-3 subscale scores before and after the cine-VR training program to assess changes in cultural self-efficacy and diabetes attitudes. In addition, we determined effect sizes using Cohen $d$ by calculating the mean difference between the pre- and postassessment responses divided by the pooled standard deviation. Finally, we calculated mean change scores for TSET-MHP subscales and DAS-3 subscales. Then, we conducted Pearson correlations with the mean change scores for each subscale and the mean subscale scores of the Presence Questionnaire. We defined statistical significance as a $P$ value less than .05 and conducted analyses in SPSS Statistics for Windows, version 26.0 (IBM Corp).

**Results**

**Overview**

A total of 76 individuals consented to participate in the study; however, 7 participants did not complete postsurveys. The final sample included 69 participants out of 76 (91% completion rate). The mean age of participants was 42.2 years (SD 13.7), 86% (59/69) identified as female, 83% (57/69) identified as White, 25% (17/69) were nurses, and 86% (59/69) were health care providers (see Table 1). Among health care providers, 72% (36/50) served more than 30% of patients with limited income and resources (ie, Medicaid) in their practice. The majority of providers cared for adult Medicaid patients (44/47, 94%), followed by 77% (30/39) who cared for older adults with Medicaid, and 69% (24/35) who cared for children with Medicaid.
Table 1. Participant demographic characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants (N=69)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>42.2 (13.7)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>59 (86)</td>
</tr>
<tr>
<td>Male</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Asian Indian</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Black</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Chinese</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Other Asian</td>
<td>2 (3)</td>
</tr>
<tr>
<td>White (non-Hispanic)</td>
<td>57 (83)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
</tr>
<tr>
<td>Community health worker</td>
<td>16 (23)</td>
</tr>
<tr>
<td>Dentist</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Dietitian</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Exercise physiologist</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Health care administrator or staff</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Nurse</td>
<td>17 (25)</td>
</tr>
<tr>
<td>Physician</td>
<td>12 (17)</td>
</tr>
<tr>
<td>Nurse practitioner</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Physical therapist</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Years in health care, n (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>7 (10)</td>
</tr>
<tr>
<td>1-5</td>
<td>15 (22)</td>
</tr>
<tr>
<td>6-10</td>
<td>6 (9)</td>
</tr>
<tr>
<td>11-15</td>
<td>3 (4)</td>
</tr>
<tr>
<td>16-20</td>
<td>5 (7)</td>
</tr>
<tr>
<td>21-25</td>
<td>14 (20)</td>
</tr>
<tr>
<td>26-30</td>
<td>4 (6)</td>
</tr>
<tr>
<td>≥31</td>
<td>5 (7)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Health care sector, n (%)</td>
<td></td>
</tr>
<tr>
<td>Health care system–affiliated clinic</td>
<td>15 (22)</td>
</tr>
<tr>
<td>Hospital</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Private practice</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Federally Qualified Health Center</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Other</td>
<td>42 (61)</td>
</tr>
<tr>
<td>Percentage of Medicaid patients served (n=50(^b)), n (%)</td>
<td></td>
</tr>
<tr>
<td>≤30%</td>
<td>9 (18)</td>
</tr>
<tr>
<td>&gt;30%</td>
<td>36 (72)</td>
</tr>
</tbody>
</table>
Participants (N=69)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants (N=69)</th>
</tr>
</thead>
<tbody>
<tr>
<td>My practice does not see Medicaid patients</td>
<td>5 (10)</td>
</tr>
<tr>
<td><strong>Age group of Medicaid patients, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Children (n=35 providers)</td>
<td>24 (69)</td>
</tr>
<tr>
<td>Adults (n=47 providers)</td>
<td>44 (94)</td>
</tr>
<tr>
<td>Older adults (n=39 providers)</td>
<td>30 (77)</td>
</tr>
</tbody>
</table>

There were 9 values missing for percentage of Medicaid patients served among the 59 providers.

**Cultural Self-Efficacy**

Mean subscale scores for the TSET-MHP are presented in Table 2. Pretraining mean scores showed that the participants had the most confidence in their Affective cultural self-efficacy (mean 8.09, SD 1.19). Prior to the training, cultural self-efficacy scores did not differ by age, gender, race, provider status, or percent of Medicaid patients.

As hypothesized, we observed positive improvements in all three of the cultural self-efficacy subscales (see Table 2): Cognitive (mean change –1.29; t\textsubscript{65} =–9.309; P <.001), Practical (mean change –1.85; t\textsubscript{65} =–9.319; P <.001), and Affective (mean change –0.75; t\textsubscript{65} =–7.067; P <.001). We observed the largest magnitude of change with the Practical subscale, with a Cohen d of 1.16 indicating a very large effect. Following the training program, the cultural self-efficacy subscale scores did not differ by age, gender, race, provider status, or percent of Medicaid patients, except for postassessment Cognitive scores. Participants who self-identified as non-White reported greater increases than White participants in postassessment Cognitive subscale scores (mean difference –0.8447; t\textsubscript{65} =–2.021; P =.047).

**Table 2.** Mean differences between Transcultural Self-Efficacy Tool–Multidisciplinary Healthcare Provider (TSET-MHP) subscale scores before and after the training program.

<table>
<thead>
<tr>
<th>TSET-MHP subscale</th>
<th>Presurvey score\textsuperscript{a}, mean (SD)</th>
<th>Postsurvey score\textsuperscript{a}, mean (SD)</th>
<th>P value</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive (n=66)</td>
<td>6.77 (1.63)</td>
<td>8.06 (1.30)</td>
<td>&lt;.001</td>
<td>0.87</td>
</tr>
<tr>
<td>Practical (n=66)</td>
<td>6.15 (1.78)</td>
<td>8.00 (1.38)</td>
<td>&lt;.001</td>
<td>1.16</td>
</tr>
<tr>
<td>Affective (n=67)</td>
<td>8.09 (1.19)</td>
<td>8.82 (1.05)</td>
<td>&lt;.001</td>
<td>0.66</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Items are rated on a 10-point scale, ranging from 1 (not confident) to 10 (totally confident).

**Diabetes Attitudes**

Mean scores for the five DAS-3 subscales are presented in Table 3. Pretraining mean scores showed that participants generally agreed with the Need for special training (mean 4.59, SD 0.38), the Seriousness of type 2 diabetes (mean 4.23, SD 0.49), the Value of tight glucose control (mean 4.10, SD 0.40), the Psychosocial impact of diabetes (mean 4.43, SD 0.43), and the Attitude toward patient autonomy (mean 4.09, SD 0.46). No differences were observed in diabetes attitudes based on age, gender, race, provider status, or percent of Medicaid patients pretraining.

As hypothesized, we observed positive improvements in all five of the diabetes attitude subscales (see Table 3): Need for special training (mean change –0.21; t\textsubscript{67} =–6.154; P <.001), Seriousness of type 2 diabetes (mean change –0.34; t\textsubscript{67} =–8.114; P <.001), Value of tight glucose control (mean change –0.13; t\textsubscript{67} =–3.029; P =.001), Psychosocial impact of diabetes (mean change –0.33; t\textsubscript{67} =–6.610; P <.001), and Attitude toward patient autonomy (mean change –0.17; t\textsubscript{67} =–3.889; P <.001). We observed the largest magnitude of change with the Psychosocial impact of diabetes subscale, with a Cohen d of 0.87 indicating a large effect. Similar to the pretraining assessment, diabetes attitudes did not differ based on age, gender, race, provider status, or percent of Medicaid patients posttraining.

**Table 3.** Mean differences between Diabetes Attitude Scale-3 (DAS-3) subscale scores before and after the training program (n=68).

<table>
<thead>
<tr>
<th>DAS-3 subscale</th>
<th>Presurvey score\textsuperscript{a}, mean (SD)</th>
<th>Postsurvey score\textsuperscript{a}, mean (SD)</th>
<th>P value</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for special training</td>
<td>4.59 (0.38)</td>
<td>4.81 (0.27)</td>
<td>&lt;.001</td>
<td>0.65</td>
</tr>
<tr>
<td>Seriousness of type 2 diabetes</td>
<td>4.23 (0.49)</td>
<td>4.57 (0.39)</td>
<td>&lt;.001</td>
<td>0.78</td>
</tr>
<tr>
<td>Value of tight glucose control</td>
<td>4.10 (0.40)</td>
<td>4.24 (0.43)</td>
<td>.001</td>
<td>0.32</td>
</tr>
<tr>
<td>Psychosocial impact of diabetes</td>
<td>4.43 (0.43)</td>
<td>4.75 (0.31)</td>
<td>&lt;.001</td>
<td>0.87</td>
</tr>
<tr>
<td>Attitude toward patient autonomy</td>
<td>4.09 (0.46)</td>
<td>4.26 (0.48)</td>
<td>&lt;.001</td>
<td>0.38</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Items are rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).
Presence in Cinematic Virtual Reality

Following the training program, we observed mean scores greater than or equal to 5.9, out of a maximum score of 7, for all five subscales: Involvement (mean 6.22, SD 0.59), Sensory Fidelity (mean 5.90, SD 0.81), Adaptation and Immersion (mean 6.22, SD 0.61), Interface Quality (mean 5.92, SD 1.31), and Presence (mean 6.28, SD 0.70). The high subscale scores demonstrate favorable perceptions of the technology and strength of presence in the cine-VR simulations. Presence in subscale scores did not differ based on age, gender, race, provider status, or percent of Medicaid patients.

Posttraining, change scores in cultural self-efficacy and diabetes attitudes were correlated with the mean subscale scores of presence. We observed only one significant correlation between the change score in Affective self-efficacy and the Interface Quality subscale score ($r=2.85$, $P=0.03$). No other significant correlations were observed between presence in cine-VR subscales and cultural self-efficacy subscale scores or diabetes attitude subscale scores (see Multimedia Appendix 1). These findings did not support the hypotheses that stated that increased presence in cine-VR would be associated with positive changes in cultural self-efficacy subscales and diabetes attitude subscales.

Discussion

Principal Findings

In this pilot study, we assessed health care providers’ and administrators’ cultural self-efficacy and diabetes attitudes before and after a 360-degree cine-VR training program. Following the training program, we observed statistically significant improvements in all three cultural self-efficacy subscales: (1) Cognitive, (2) Practical, and (3) Affective. The largest magnitude of effect was observed with the Practical subscale, which corresponds to confidence in interviewing patients about social determinants of health. In addition, all five diabetes attitude subscales improved significantly posttraining: (1) Need for special training, (2) Seriousness of type 2 diabetes, (3) Value of tight glucose control, (4) Psychosocial impact of diabetes, and (5) Attitude toward patient autonomy, with the largest magnitude of change observed in Psychosocial impact of diabetes. Lastly, we observed high scores for presence in cine-VR, indicating favorable perceptions of the technology and immersion in the 360-degree virtual environment. Contrary to expectations, only one positive change score in Affective self-efficacy was correlated with increased presence in cine-VR.

Comparison With Prior Work

Effective cine-VR simulations provide a platform to practice and acquire skills that will later translate to clinical outcomes concerning patient care; in addition, they afford participants the opportunity to practice clinical judgment and apply problem-solving skills in a risk-free, replicable clinical environment [29,30]. Cine-VR technology offers new opportunities for clinical assessment and intervention. Advances in virtual reality technologies can now support the creation of low-cost, yet sophisticated, immersive simulations, capable of running on consumer-level computing devices [31]. Compared to traditional video training, the immersive qualities of cine-VR affect the participant’s ability to more strongly retrieve the experience from memory, suggesting that cine-VR experiences become part of an autobiographical associative network, whereas a conventional video experience remains an isolated episodic event [32].

Existing research in narrative health promotion demonstrates the power of culturally tailored stories as engaging content to positively affect attitudes, beliefs, and behaviors. Qualitative results show that the digital storytelling more positively affects participants than traditional face-to-face training on its own, specifically in four growth areas: truth-telling, sense-making, social support, and feeling valued [33]. Research concerning digital storytelling and its uses within health care are only in their infancy in terms of discovering applications and uses. However, recent studies demonstrate that digital stories allow for a deeper understanding of an experience rather than simply hearing an explanation of that experience [34]. Our research supports this finding. Our findings suggest that this innovative cine-VR program can be used to educate providers about type 2 diabetes, social determinants of health, and Appalachian culture, which, in turn, may enhance the delivery of high-quality, evidence-based diabetes care in rural Appalachian Ohio. Additional research is needed to determine the impact of the training on patient care and health outcomes.

Finally, presence describes the extent to which a participant feels present or immersed in a virtual environment [35,36] and is commonly regarded as a necessary mediator that allows real emotions to be activated [37,38]. We hypothesized that higher levels of presence would be associated with positive changes in cultural self-efficacy and diabetes attitudes. We observed only one significant correlation between the change score in Affective self-efficacy and the Interface Quality subscale score. This finding suggests that participants who felt less distracted by the headset or experienced fewer delays with the simulations showed a greater improvement in the Affective self-efficacy scores posttraining. We observed no other significant correlations between positive change scores and presence. This may be explained by the limited variability in presence subscale scores and the overall high level of presence measured in the study. The strength of this 360-degree cine-VR simulation training program is the realism afforded by providing the participant access to the whole environment as compared to traditional virtual reality (eg, animated environments and characters), which has been criticized as being too unrealistic [39].

Limitations

Limitations of this study include the small homogeneous sample, selection bias, social desirability bias, and lack of a control group. While a final sample of 69 participants is small, our a priori power analysis determined that a sample size of 34 paired participants was sufficient to achieve 80% power and a level of significance of $P<.05$. We successfully doubled the required sample size estimate. However, data from 69 providers and administrators from one geographic region limits the generalizability of the findings to other providers. Further, the predominantly White study sample limits the generalizability to all providers; however, the racial and ethnic distribution of
the study sample (83% White) is reflective of the racial and ethnic distribution in southeastern Ohio (95% White) [40]. Next, our findings may be susceptible to selection bias, as individuals who volunteered to participate may have been more willing or motivated to participate in a novel educational program about diabetes, social determinants of health, and Appalachian culture. In addition, the responses may be susceptible to selection bias given the participants may have felt undue pressure to provide positive feedback on the training session. A similar susceptibility to selection bias may be prescribed to the use of new technology encouraging people to provide positive feedback. Finally, this study presents findings from a 3-hour cine-VR training program on type 2 diabetes in rural Appalachia. We did not include a control condition as a comparison group. Future research should use a randomized controlled design to assess the impact of two different educational interventions on providers’ and administrators’ cultural self-efficacy and diabetes attitudes.

Conclusions

Continuing medical education is an important component of clinical care for all providers. Health care providers and administrators need ongoing and repeated training to help them improve and maintain their knowledge, stay current with the latest developments, address real-world challenges, and learn effective team management skills. Our findings support the notion that 360-degree cine-VR education is an innovative approach to improve cultural self-efficacy and diabetes attitudes among health care providers and administrators. The long-term impact of cine-VR education on cultural self-efficacy and diabetes attitudes needs to be determined.

Acknowledgments

This study was part of the Medicaid Equity Simulation Project funded by the Ohio Department of Medicaid and administered by the Ohio Colleges of Medicine Government Resource Center. The views expressed in this publication about the cine-VR simulations are solely those of the creators and do not represent the views of the state of Ohio or federal Medicaid programs.

Conflicts of Interest
None declared.

Multimedia Appendix 1

Correlations among subscale scores of presence in virtual reality, change scores in cultural self-efficacy, and Diabetes Attitude Scale-3 (DAS-3) subscales (n=65).

References


Abbreviations

cine-VR: virtual reality cinema
DAS-3: Diabetes Attitude Scale-3
TSET-MHP: Transcultural Self-Efficacy Tool–Multidisciplinary Healthcare Provider

©Elizabeth Ann Beverly, Carrie Love, Matthew Love, Eric Williams, John Bowditch. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 27.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Feasibility of the Web-Based Intervention Designed to Educate and Improve Adherence Through Learning to Use Continuous Glucose Monitor (IDEAL CGM) Training and Follow-Up Support Intervention: Randomized Controlled Pilot Study

Madison B Smith¹, PhD, RN, CDCES; Anastasia Albanese-O’Neill², PhD; Yingwei Yao³, PhD; Diana J Wilkie³, PhD, RN, FAAN; Michael J Haller², MD; Gail M Keenan⁴, PhD, RN, FAAN

¹College of Nursing, University of Florida, Gainesville, FL, United States
²Department of Pediatrics, College of Medicine, University of Florida, Gainesville, FL, United States
³Department of Biobehavioral Nursing Science, College of Nursing, University of Florida, Gainesville, FL, United States
⁴Department of Family, Community and Health Systems Science, College of Nursing, University of Florida, Gainesville, FL, United States

Corresponding Author:
Madison B Smith, PhD, RN, CDCES
College of Nursing
University of Florida
1225 Center Drive, HPNP Room 3229, PO Box 100197
Gainesville, FL
United States
Phone: 1 4074432555
Email: madisonbricksmith@gmail.com

Abstract

Background: Proper training and follow-up for patients new to continuous glucose monitor (CGM) use are required to maintain adherence and achieve diabetes-related outcomes. However, CGM training is hampered by the lack of evidence-based standards and poor reimbursement. We hypothesized that web-based CGM training and education would be effective and could be provided with minimal burden to the health care team.

Objective: The aim of this study was to perform a pilot feasibility study testing a theory-driven, web-based intervention designed to provide extended training and follow-up support to adolescents and young adults newly implementing CGM and to describe CGM adherence, glycemic control, and CGM-specific psychosocial measures before and after the intervention.

Methods: The “Intervention Designed to Educate and improve Adherence through Learning to use CGM (IDEAL CGM)” web-based training intervention was based on supporting literature and theoretical concepts adapted from the health belief model and social cognitive theory. Patients new to CGM, who were aged 15-24 years with type 1 diabetes for more than 6 months were recruited from within a public university’s endocrinology clinic. Participants were randomized to enhanced standard care or enhanced standard care plus the IDEAL CGM intervention using a 1:3 randomization scheme. Hemoglobin A₁c levels and psychosocial measures were assessed at baseline and 3 months after start of the intervention.

Results: Ten eligible subjects were approached for recruitment and 8 were randomized. Within the IDEAL CGM group, 4 of the 6 participants received exposure to the web-based training. Half of the participants completed at least 5 of the 7 modules; however, dosage of the intervention and level of engagement varied widely among the participants. This study provided proof of concept for use of a web-based intervention to deliver follow-up CGM training and support. However, revisions to the intervention are needed in order to improve engagement and determine feasibility.

Conclusions: This pilot study underscores the importance of continued research efforts to optimize the use of web-based intervention tools for their potential to improve adherence and glycemic control and the psychosocial impact of the use of diabetes technologies without adding significant burden to the health care team. Enhancements should be made to the intervention to increase engagement, maximize responsiveness, and ensure attainment of the skills necessary to achieve consistent use and improvements in glycemic control prior to the design of a larger well-powered clinical trial to establish feasibility.

KEYWORDS
type 1 diabetes mellitus; continuous glucose monitor; web-based training; diabetes education; intervention

Introduction
Background
Historically, adolescents and young adults have demonstrated the poorest glycemic control compared to younger children and older adults; yet, they remain the most resistant to adopting newly developed technologies that could significantly improve type 1 diabetes (T1D) outcomes [1]. The continuous glucose monitor (CGM) can substantially improve glycemic control when worn consistently [2-4]. Despite the recognized benefit, only 24% of the adolescents and 22% of the young adults with T1D are current CGM users compared to 51% and 37% of children (aged less than 6 years and 6-12 years, respectively) and 37% and 34% of the adults (aged 26-50 years and older than 50 years, respectively) [1]. Even fewer adolescents and young adults wear the device with the consistency associated with improved glycemic control [3,5]. To foster adherence to the device and improve outcomes, experts cite the importance of training and follow-up support during the first few months to ensure proper use of CGMs [6]. Thus, a pilot randomized controlled trial was implemented to evaluate the feasibility of the web-based “Intervention Designed to Educate and improve Adherence through Learning to use CGM” or the IDEAL CGM.

CGM Use
An international consensus statement released by key leaders regarding the use of CGM in children and adolescents stated that proper training is necessary for patients to use CGM correctly [6]. Recommendations include maintaining a high level of contact with families during the first few months of wear, which incorporates start-up training and realistic expectation setting, in addition to follow-up visits after CGM implementation to download data, review alarm settings, encourage ongoing CGM use, and address potential barriers to use [6]. These efforts take a significant amount of time and health care resources without financial reimbursement available to offset costs [7]. CGM education does not yet have established standards that are widely recognized and there is little evidence available to link educational efforts to diabetes-related outcomes [7-9].

The study of human factors works to leverage the characteristics and limitations of human interactions to improve the design of systems and use of technology [10]. Psychosocial factors play a significant role in patient acceptance and use of these technologies [11]. These factors include satisfaction (hassles and benefits of use) [12-15], self-efficacy [16], quality of life [13,17,18], and emotional distress [12]. Interventions targeting human factors related to CGM use represent an opportunity to improve adherence rates and patient-reported outcomes [12]. The association between human factors and consistent use suggests that clinical interventions targeting these modifiable factors could have an effect on CGM; however, such interventions have yet to be studied [11].

Study Intervention Rationale
Patients desire access to diabetes care that is flexible and adaptive to their individual needs in regard to timing, frequency, and form of contact [19], especially when knowledge deficiencies arise [20]. Over 96% of the young adults have been reported to seek further diabetes education outside of clinic with 81% referring to websites and 30% using web-based chat rooms and blogs [20]. The widespread acceptance of web-based resources by this population supports the use of mobile-based and web-based programs to provide tailored education to adolescents and young adult patients with T1D [21-28], without increasing the health care burden related to increased training and follow-up needs. This pilot study aimed to evaluate the feasibility of delivering a theory-driven, web-based intervention to provide follow-up training and peer support to adolescents and young adults new to CGM and to describe diabetes-related outcomes before and after the interventional period.

Methods
Design and Setting
Using a randomized control-group pretest-posttest design, we recruited 8 participants from a large public university’s pediatric endocrinology clinic between March 2018 and July 2018 during routine office visits and scheduled CGM trainings in clinic. Participants were randomized to enhanced standard care or enhanced standard care plus the intervention by using a 1:3 allocation scheme. This study was approved as expedited minimal risk by the University of Florida Institutional Review Board.

Subjects
The inclusion criteria were as follows: (1) ability to read and speak English; (2) diagnosed with T1D for at least 6 months; (3) aged between 15 and 24 years at the time of enrollment; (4) access to a smartphone, tablet, or laptop/desktop computer with high speed internet access and speaker; and (5) intended use of a Dexcom G5 CGM. Participants were required to be new to CGM or have no previous CGM use within the last 3 months. Participants with significant learning disabilities or inability to comply with the study protocol were excluded. Eligible subjects were identified via a review of upcoming medical appointments, which indicated patients scheduled for CGM training. Recruitment of subjects occurred on a rolling basis within the clinical setting.

Procedure
All participants received at least one 60-minute, face-to-face, basic CGM education and training session conducted by the regular clinical team. This training was considered enhanced standard care and took place outside of the study, prior to recruitment and enrollment (Table 1). After obtaining consent and assent (for participants aged 17 years or younger), baseline hemoglobin A1C (HbA1c) measures were collected. A 1-week
CGM run-in period was completed prior to baseline questionnaires. The web-based training intervention was delivered over a 6-week period. Adherence and glycemic control outcomes were assessed at 3 months from the baseline.

Allocation to the intervention took place using sealed envelopes generated by the investigators to reveal randomization status. Participants within the enhanced standard care group followed an identical study activity timeline, with the exception of exposure to the IDEAL CGM web-based training program. No participant was restricted from accessing additional CGM educational materials or device support throughout the study. Participants were compensated up to US $50 for completion of the initial and follow-up surveys and HbA₁c measures; compensation was not dependent on completion of the intervention or adherence to CGM.

Table 1. Study activity timeline demonstrating activities over the 3-month study period.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Week –1</th>
<th>Week 0</th>
<th>Weeks 1-6</th>
<th>Week 7</th>
<th>Weeks 11-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced standard CGM training(^b)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study recruitment</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surveys/tools(^c)</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Introduction module(^d)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Web-based intervention(^d)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Exit satisfaction survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Hemoglobin A₁c measures(^d)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Download CGM data(^c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

\(^a\)CGM: continuous glucose monitor.
\(^b\)Standardized training completed per clinic’s enhanced standard care, prior to enrollment in study.
\(^c\)Includes continuous glucose monitor self-efficacy survey, satisfaction scale surveys, and knowledge assessment tool.
\(^d\)Indicates activity only designated for the intervention arm.
\(^e\)Objective measure of continuous glucose monitor adherence over the 3-month study period.

IDEAL CGM Web-Based Intervention

Human factors or individual beliefs associated with adherence to CGM (ie, benefits, hassles, self-efficacy) [11] are well known concepts supported by the health belief model and social cognitive theory [29,30]. The model, shown in Figure 1, used constructs of behavior change and learning theories to provide follow-up CGM training and social support to overcome perceived hassles related to CGM use and encourage behaviors that influence expected outcomes. Further, action-oriented learning strategies, seen in Table 2 [31-42], were incorporated into the IDEAL CGM intervention to create a dynamic learning process that motivated participation and skill attainment.
Figure 1. A conceptual model to support the design of the intervention and determined outcome measures. CGM: continuous glucose monitor; HbA1c: hemoglobin A1c; CGM-SE: CGM self-efficacy; CGM-SAT: CGM-satisfaction scale.

Table 2. Evidence to support action-oriented learning strategies incorporated into the web-based intervention design.

<table>
<thead>
<tr>
<th>Action-oriented learning strategy</th>
<th>Component of intervention</th>
<th>Literature to support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setting</td>
<td>Personal goal setting</td>
<td>1 of the 3 main factors to affect likelihood a person will change a health behavior [31]</td>
</tr>
<tr>
<td>Outcome expectancies: result an individual anticipates from taking action [31]</td>
<td>CGM expectation setting</td>
<td>1 of the 3 main factors to affect likelihood a person will change a health behavior [31]. Failure to meet expectations is one of the top cited reasons for poor CGM adherence [12,15,32-36]. Realistic expectations while using CGM were associated with better glycemic control and patient success [37]</td>
</tr>
<tr>
<td>Behavioral capabilities: knowledge and skill to perform given behavior [31]</td>
<td>Knowledge acquisition through provided materials</td>
<td>Proper training is necessary for patients to use CGM correctly [6]. Difficult to use technology is one of the top cited reasons for poor CGM adherence [12,15,32-36]</td>
</tr>
<tr>
<td>Cues to action: factors that promote action [31]</td>
<td>Push notifications and email reminders to access LMSb</td>
<td>Reminders to access and utilize web-based programs were critical to previously tested web-based intervention’s success [22,26,38,39]</td>
</tr>
<tr>
<td>Monitoring progress [31], Reinforcing learned behaviors [31]</td>
<td>Knowledge assessment checks</td>
<td>Patients who consistently applied themselves to homework assignments, worksheets, and brief quizzes to reinforce learning and evaluate information gaps were observed to be most successful with SAPc [9]</td>
</tr>
<tr>
<td>Observational learning (modeling): learning through the experience of credible others rather than through their own experiences [31]</td>
<td>Discussion boards with peers (content monitored by health care professionals)</td>
<td>Discussion boards were highly utilized when incorporated into program designs [22,40]. Young adults utilize web-based resources, websites, discussion boards, and blogs to augment peer and family support [41,42]. Peer-led education provided an opportunity to learn real-life explanations for problems not addressed in clinic-based learning [20]</td>
</tr>
</tbody>
</table>

aCGM: continuous glucose monitor.  
bLMS: learning management system.  
cSAP: sensor-augmented pump therapy.

The IDEAL CGM program was delivered via a learning management system that required a personal login and password to access via the desktop or mobile phone [43]. See Figure 2 for screenshots of the web-based and mobile-based home pages of the IDEAL CGM platform, which included access to asynchronous educational modules designed using professionally
supported educational topics and training materials. Topics were created based on top patient-reported hassles leading to inconsistent or discontinued CGM use (ie, unmet expectations, alarm fatigue, placement/adhesion issues) [12], as well as training concepts pertinent to developing CGM self-efficacy and underscoring the benefits of use (ie, guidelines for treatment decisions, uploading/sharing data, and interpreting data; Multimedia Appendix 1). Peer-led discussion boards were linked to each module, which were intended to establish social support while facilitating peer-led observational learning. A health care professional monitored the discussion boards for appropriateness of content and provided tailored responses. Each module was designed using the same format and included a summary of the module topic, a “to-do” list with actionable items, a list of learning objectives, links to recorded video materials, additional materials to review, and recommended resources. Each week, proposed tasks included the review of recorded video materials, written educational content, and visual imagery, completion of the knowledge assessment checks, and participation within the peer-led discussion boards.

**Figure 2.** Screenshot of the IDEAL CGM (Intervention Designed to Educate and improve Adherence through Learning to use continuous glucose monitor) homepage. A. web-based and B. mobile-based.

**Study Measures**

We intended to examine the acceptability of the protocol, intervention dosage, participant responsiveness (user engagement in knowledge checks and discussion boards), and patient satisfaction with the IDEAL CGM program. Diabetes-related measures were described before and after the intervention and in relation to dosage of the intervention. Study
data and survey responses were collected and managed using institutional review board–approved Research Electronic Data Capture (REDCap) tools hosted at the University of Florida [44]. REDCap is a secure, web-based app designed to support data capture for research studies. Electronic medical records and joint parent-youth interviews provided demographic and clinical data.

Feasibility Measures

Acceptability of the Protocol

Measures included recruitment and retention with a goal of at least 80% completion of baseline and follow-up measures.

Dosage and Participant Responsiveness

The learning management system collected and stored individual data related to dosage (ie, time spent, number of views, type of views) and participant responsiveness (ie, knowledge check submissions and discussion board posts) within the IDEAL CGM intervention.

Exit Satisfaction Survey

The exit satisfaction survey included 16 questions from the validated Flashlight Current Student Inventory, which was designed to gather information about a participant’s reaction to various teaching and learning practices [45]. The exit satisfaction survey used a 5-point Likert scale and open-ended questions to assess satisfaction related to the CGM training provided. Higher scores indicate more favorable satisfaction levels. The overall score is the mean of the item scores.

Diabetes-Related Measures

CGM Adherence

Usage data were collected by the CGM receiver and manually downloaded or automatically synced to a diabetes management platform. Adherence is described as the percentage of days that the CGM was worn over a 90-day period, with target adherence rates set to greater than 85%.

Glycemic Control

HbA1c levels were measured using a DCA Vantage Analyzer (Siemens).

CGM Satisfaction

The CGM Satisfaction Scale [46], a 44-item validated measure, uses a 5-point Likert scale to assess satisfaction specific to CGM use and includes 2 subscales of “lack of hassles” and “benefits.” Higher scores indicate a more favorable impact and satisfaction with CGM use. Overall score is the mean of item scores.

CGM Self-efficacy

The CGM self-efficacy [16] version for youth older than 13 years, which is a 15-item validated measure, uses a 7-point Likert scale to assess the confidence of youth and parents to manage the technical and behavioral aspects of CGM use. Scores range from 0 to 100. CGM self-efficacy scores greater than 80 are considered “high” and are associated with adherence to CGM use and lower HbA1c levels after 3 months [16]. The CGM self-efficacy survey has not yet been validated in youth 18 years or older.

Knowledge Assessment

The 20-question unvalidated assessment designed for the study used a multiple choice questionnaire to measure the attainment of knowledge related to the key aspects of CGM use. The knowledge assessment was scored as 0%-100%.

Data Analysis

Intention-to-treat analysis was performed based on the randomization status of each participant. Participants randomized to the intervention group were included within analysis, regardless of the actual dosage or participant responsiveness within the intervention. Analysis was performed in SPSS (Version 25, IBM Corp). Descriptive statistics were presented for individual participant data with group median and range provided.

Results

Measures of Feasibility

Acceptability of the Protocol

The acceptability of the protocol is demonstrated by the study flow diagram (Figure 3). Of the 10 patients assessed for eligibility, 8 (80%) agreed to participate and were randomized to the enhanced standard care versus intervention plus enhanced standard care groups. For ease of interpreting study results, participants (P) were numbered 1-8 and were categorized based on intervention (i) or enhanced standard care/control group (c). P1-i through P6-i identify those randomized to the intervention, while P7-c and P8-c were randomized to the enhanced standard care group. The baseline and clinical characteristics of the 2 groups were comparable, as shown in Table 3.

This study demonstrated the ability to retain participants with a very low attrition rate. All survey measures were completed. Six of the 8 participants (75%) returned to clinic within the 3-month (SD, 2 weeks) study window for HbA1c assessment, while the assessments for the other 2 participants (P1-i and P4-i) were performed outside of the intended window. CGM data were collected from 7 participants (88%) at follow-up. P1-i failed to bring the personal receiver in for upload and was unable to upload remotely.
Figure 3. Study flow diagram. CGM: continuous glucose monitor.
Table 3. Baseline characteristics and clinical features of the enrolled participants.

<table>
<thead>
<tr>
<th>Participant (P)</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Race</th>
<th>Ethnicity</th>
<th>Current pump use</th>
<th>Previous CGM&lt;sup&gt;a&lt;/sup&gt; use</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-i</td>
<td>17</td>
<td>Male</td>
<td>White</td>
<td>Non-Hispanic</td>
<td>Yes</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>P2-i</td>
<td>16</td>
<td>Female</td>
<td>Mixed</td>
<td>Non-Hispanic</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>P3-i</td>
<td>17</td>
<td>Male</td>
<td>White</td>
<td>Non-Hispanic</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>P4-i</td>
<td>15</td>
<td>Female</td>
<td>White</td>
<td>Non-Hispanic</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>P5-i</td>
<td>20</td>
<td>Female</td>
<td>White</td>
<td>Hispanic</td>
<td>No</td>
<td>Brand: Dexcom</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Duration of use: 2 weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Date: 2 years prior</td>
</tr>
<tr>
<td>P6-i</td>
<td>16</td>
<td>Male</td>
<td>White</td>
<td>Non-Hispanic</td>
<td>No</td>
<td>Brand: Dexcom</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Duration: 12 weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Date: 6 months prior</td>
</tr>
<tr>
<td>P7-c</td>
<td>17</td>
<td>Female</td>
<td>White</td>
<td>Non-Hispanic</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>P8-c</td>
<td>18</td>
<td>Male</td>
<td>Not reported</td>
<td>Hispanic</td>
<td>Yes</td>
<td>Brand: Medtronic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Duration of use: 1 week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Date: 4-5 years prior</td>
</tr>
</tbody>
</table>

<sup>a</sup>CJM: continuous glucose monitor.

<sup>b</sup>N/A: not applicable (they were naïve to CGM prior to study).

**Dosage and Participant Responsiveness**

The number of modules viewed by the participants varied widely. The overall average view rate of the modules was 48% (3.3/7 modules). In total, 4 of the 6 intervention participants completed the steps required to login to the IDEAL CGM program and view the training modules; the remaining 2 never logged into the intervention platform. Half of the intervention participants (n=3) were engaged in at least 5 of the 7 modules or more than 70% of the intended modules. However, the time spent within the modules and participant responsiveness varied. The median time spent within the web-based platform was 32 minutes (range 0-138 minutes). Figure 4 displays the dosage and type of engagement within the web-based intervention for each participant. P2-i and P3-i completed specific knowledge checks more than once (range 2-5 times). See Multimedia Appendix 2 for additional details regarding the frequency and type of participant engagement within each module.
Participant Satisfaction

Overall, participants within both groups reported being satisfied with their CGM training and perceived level of active and collaborative learning. Four participants within the intervention group indicated they were “very satisfied” with their CGM education, while 2 were “satisfied” (P4-i and P6-i). One participant within the standard care group reported being “very satisfied” while one reported being “satisfied.” Scores ranged from 3.3 to 4.4 within the intervention group and 2.9 to 3.0 within the enhanced standard care group.

When asked to describe what they liked most about the CGM training provided, participants from the intervention group reported “being able to relate to other peers,” “the people were relatable to my lifestyle and how to accommodate any problems I had,” and “they made it easy to understand and easy to use for me.” Only participants with exposure to the intervention included comments related to peer engagement and observational learning. When asked to describe what they disliked the most, participants from the intervention group reported the need for “more study reminders,” the use of “shorter videos,” and the need to “rewatch the videos.” A complete list of open-ended participant feedback regarding CGM training is included in Multimedia Appendix 3.

Diabetes-Related Outcomes

Participant data are summarized in Table 4.
Table 4. Diabetes-related outcome measures at baseline and follow-up per participant.

<table>
<thead>
<tr>
<th>Measures</th>
<th>P1-i(^a)</th>
<th>P2-i(^a)</th>
<th>P3-i(^a)</th>
<th>P4-i(^a)</th>
<th>P5-i(^a)</th>
<th>P6-i(^a)</th>
<th>P7-c(^b)</th>
<th>P8-c(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGM(^c) adherence (%) 3 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>11.6</td>
<td>&gt;14</td>
<td>12.3</td>
<td>10.2</td>
<td>8.5</td>
<td>&gt;14</td>
<td>8.7</td>
<td>10.7</td>
</tr>
<tr>
<td>Follow-up</td>
<td>9.8</td>
<td>&gt;14</td>
<td>9.8</td>
<td>9</td>
<td>8.4</td>
<td>&gt;14</td>
<td>9.3</td>
<td>9.5</td>
</tr>
<tr>
<td>Glycemic control (HbA(^1c) %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>4.7</td>
<td>1.3</td>
<td>3.8</td>
<td>3.6</td>
<td>4.3</td>
<td>3.5</td>
<td>3.9</td>
<td>4.2</td>
</tr>
<tr>
<td>Follow-up</td>
<td>3.9</td>
<td>4.0</td>
<td>3.9</td>
<td>3.8</td>
<td>4.3</td>
<td>3.6</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>CGM satisfaction survey score (max score 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>100</td>
<td>100</td>
<td>94</td>
<td>83</td>
<td>97</td>
<td>68</td>
<td>93</td>
<td>96</td>
</tr>
<tr>
<td>Follow-up</td>
<td>89</td>
<td>84</td>
<td>92</td>
<td>78</td>
<td>99</td>
<td>50</td>
<td>98</td>
<td>84</td>
</tr>
<tr>
<td>CGM self-efficacy survey score (max score 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>40</td>
<td>65</td>
<td>80</td>
<td>65</td>
<td>70</td>
<td>40</td>
<td>60</td>
<td>85</td>
</tr>
<tr>
<td>Follow-up</td>
<td>55</td>
<td>80</td>
<td>65</td>
<td>60</td>
<td>55</td>
<td>45</td>
<td>70</td>
<td>85</td>
</tr>
</tbody>
</table>

\(^a\)Participant in the intervention group.
\(^b\)Participant in the enhanced standard care group.
\(^c\)CGM: continuous glucose monitor.
\(^d\)Not available.

CGM Adherence
CGM adherence was clustered around 3 levels of use for the intervention group (P1-i to P6-i). One participant reached recommended use of at least 85% (P3-i, 80/90 days, 89%); 2 participants fell just shy of recommendations with greater than 60% use (P2-i, 55/90 days, 61%; P5-i, 56/90 days, 62%), and 2 participants had less than 15% use (P4-i, 9/90 days, 10%; P6-i, 11/90 days, 12%). The 2 participants within the standard care group reached recommended use of at least 85% (P7-c, 80/90 days, 89%; P8-c, 85/90 days, 94%). No CGM adherence data were collected for participant P1-i.

Glycemic Control
Four participants within the intervention group saw an improvement in HbA\(^1c\) levels, ranging from 0.1% to 2.5%. The remaining 2 participants randomized to the intervention arm (P2-i and P6-i) had an HbA\(^1c\) level of greater than 14% at baseline and follow-up; therefore, potential improvements could not to be detected using the point-of-care HbA\(^1c\) analyzers. Of the participants within the enhanced standard care group, P8-c saw a 1.2% improvement in HbA\(^1c\) levels, while P7-c saw a worsening in HbA\(^1c\) levels (8.7% increased to 9.3%) after 3 months of CGM use.

Psychosocial Measures
Within the intervention group, median CGM satisfaction scale scores improved from 3.7 at baseline (range 1.3-4.7) to 3.9 at follow-up (range 3.6-4.3). Within the enhanced standard care group, P8-c described a -0.3 decline in satisfaction from 4.2 to 3.9 while the satisfaction of P7-c remained unchanged from baseline to follow up (3.9). Within the intervention group, the median CGM self-efficacy scores decreased from 96 at baseline (range 68-100) to 87 at follow-up (range 50-99). Within the enhanced standard care group, 1 participant (P7-c) showed an increase in the score while the other participant (P8-c) showed a decrease in the score. Despite decreases in the self-efficacy, follow-up CGM self-efficacy scores remained “high” (greater than 80) for all except for the 2 participants with the lowest CGM adherence (9/90 days, 10% and 11/90 days, 12%) and limited to no engagement within the intervention (P4-i and P6-i) [16].

Knowledge Assessment
Within the intervention group, median CGM knowledge assessment scores were 65 at baseline (range 40-80), which decreased to 58 at follow-up (range 45-80). CGM knowledge assessment scores widely varied from baseline to follow-up, with some participants demonstrating knowledge attainment while others showed worsened scores. The 2 participants with exposure to at least 6 of the intervention modules demonstrated the greatest improvements in CGM knowledge, with a 15-point increase in score.

Discussion
Principal Findings
This pilot study examined the feasibility of the IDEAL CGM intervention and described patient adherence to CGM, changes in glycemic control, psychosocial measures, and knowledge levels in the intervention and enhanced standard care groups. Initial findings from the pilot sample of 8 participants...
demonstrated proof of concept and provided key design considerations for future efforts aimed at utilizing web-based training interventions. Overall, patients were satisfied with the IDEAL CGM training intervention and perceived high levels of active and collaborative learning during CGM training. Open-ended responses suggested the impact of the peer-led discussions on perceived social support. Additional research is necessary to determine the feasibility of using web-based training to improve adherence to CGM in adolescents and young adults new to CGM use. The heterogeneity of this population suggests the vastly differing levels of training and follow-up support necessary to improve CGM adherence and help patients reach glycemic targets. Aside from training alone, this study demonstrates the importance of considering baseline characteristics, factors motivating CGM use, intervention participation, and the translation of knowledge into learned behaviors. While some participants reached clinically relevant improvements in HbA1c levels and sustained CGM use following relatively minimal to moderate levels of personalized training and follow-up support, other participants were likely in need of additional resources to maximize these outcomes. Aside from behavior, confounding variables such as diabetes distress, family conflict, perceived support, and psychological barriers should be investigated when limited improvements in HbA1c levels occur despite high CGM adherence.

**Limitations**

Study recruitment and the potential to determine feasibility were limited by the Food and Drug Administration’s approval of an upgraded version of the Dexcom CGM (Dexcom G6) ahead of the expected timeline. Both providers and patients often opt to wait until the release of the newest CGM technology. When possible, future training interventions should create materials that remain relevant, despite updates within the technology, and should exist in a format that can be easily updated to keep up with the continuous evolution and development of diabetes technology. Further, as CGM use becomes the standard of care within T1D management, many patients are started on these systems soon after diagnosis. Historically, research protocols have excluded patients recently diagnosed within the last 6-12 months to account for confounding variables affecting improvements in glycemic control (ie, intensive insulin therapy and residual beta-cell function). However, this shift within the clinical paradigm will likely affect studies’ ability to recruit patients naïve to diabetes technologies 6-12 months past diagnosis.

**Conclusion**

Web-based training and support interventions should continue to be explored for their potential to improve adherence and glycemic outcomes, while minimizing the burden or psychosocial impact of use during the uptake of new diabetes technologies. Web-based interventions increase patient exposure to diabetes-self management education with little to no added burden to the health care team. Continued efforts should work to establish evidence-based training standards and follow-up support methods necessary to achieve the diabetes-related outcomes associated with CGM use. Further research is needed to demonstrate the feasibility of using a web-based intervention to increase knowledge, maximize patient responsiveness, and ensure the successful uptake of and consistent use of CGM technology by adolescents and young adults.

**Acknowledgments**

This study was funded by the University of Florida Department of Pediatrics Children’s Miracle Network Grant. The authors would like to thank Giustina Ventura, James Kocher, and Danean Ermentrout for their contribution and support during the execution of this pilot study.

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**

Description of module topics within IDEAL CGM (Intervention Designed to Educate and Improve Adherence Through Learning to Use Continuous Glucose Monitor) training intervention.

[DOCX File, 14 KB - diabetes_v6i1e15410_app1.docx]

**Multimedia Appendix 2**

Detailed view of participant dosage and responsiveness within the IDEAL CGM (Intervention Designed to Educate and Improve Adherence Through Learning to Use Continuous Glucose Monitor) training intervention.

[PNG File, 138 KB - diabetes_v6i1e15410_app2.png]

**Multimedia Appendix 3**

Open-ended exit satisfaction survey responses from each participant.

[DOCX File, 16 KB - diabetes_v6i1e15410_app3.docx]

**Multimedia Appendix 4**

CONSORT-eHEALTH checklist (V 1.6.1).
References


24. Witt S. Glu: An online type 1 diabetes information community. SLIS Student Research Journal. 2016. URL: http://scholarworks.sjsu.edu/slisrj/vol6/iss1/3 [accessed 2020-12-17]


43. 2019. Canvas. URL: https://www.canvaslms.com [accessed 2020-12-17]


Abbreviations

CGM: continuous glucose monitor
IDEAL: Intervention Designed to Educate and Improve Adherence Through Learning
HbA₁c: hemoglobin A₁c
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

©Madison B Smith, Anastasia Albanese-O'Neill, Yingwei Yao, Diana J Wilkie, Michael J Haller, Gail M Keenan. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 09.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Evaluation of a Diabetes Remote Monitoring Program Facilitated by Connected Glucose Meters for Patients With Poorly Controlled Type 2 Diabetes: Randomized Crossover Trial

Daniel J Amante, PhD, MPH; David M Harlan, MD; Stephenie C Lemon, PhD; David D McManus, MD, MSc; Oladapo O Olaitan, MS; Sherry L Pagoto, PhD; Ben S Gerber, MD, MPH; Michael J Thompson, MD

1Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States
2Division of Diabetes, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
3Division of Cardiovascular Medicine, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
4Department of Allied Health Sciences, Institute for Collaborations on Health, Interventions, and Policy, University of Connecticut, Storrs, CT, United States
5Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States

Corresponding Author:
Daniel J Amante, PhD, MPH
Department of Population and Quantitative Health Sciences
University of Massachusetts Medical School
368 Plantation St
Worcester, MA, 01655
United States
Phone: 1 5088568480
Email: daniel.amante@umassmed.edu

Abstract

Background: Patients with poorly controlled type 2 diabetes (T2D) experience increased morbidity, increased mortality, and higher cost of care. Self-monitoring of blood glucose (SMBG) is a critical component of diabetes self-management with established diabetes outcome benefits. Technological advancements in blood glucose meters, including cellular-connected devices that automatically upload SMBG data to secure cloud-based databases, allow for improved sharing and monitoring of SMBG data. Real-time monitoring of SMBG data presents opportunities to provide timely support to patients that is responsive to abnormal SMBG recordings. Such diabetes remote monitoring programs can provide patients with poorly controlled T2D additional support needed to improve critical outcomes.

Objective: To evaluate 6 months of a diabetes remote monitoring program facilitated by cellular-connected glucose meter, access to a diabetes coach, and support responsive to abnormal blood glucose recordings greater than 400 mg/dL or below 50 mg/dL in adults with poorly controlled T2D.

Methods: Patients (N=119) receiving care at a diabetes center of excellence participated in a two-arm, 12-month randomized crossover study. The intervention included a cellular-connected glucose meter and phone-based diabetes coaching provided by Livongo Health. The coach answered questions, assisted in goal setting, and provided support in response to abnormal glucose levels. One group received the intervention for 6 months before returning to usual care (IV/UC). The other group received usual care before enrolling in the intervention (UC/IV) for 6 months. Change in hemoglobin A1c (HbA1c) was the primary outcome, and change in treatment satisfaction was the secondary outcome.

Results: Improvements in mean HbA1c were seen in both groups during the first 6 months (IV/UC −1.1%, SD 1.5 vs UC/IV −0.8%, SD 1.5; P<.001). After crossover, there was no significant change in HbA1c in IV/UC (mean HbA1c change +0.2, SD 1.7, P=.41); however, those in UC/IV showed further improvement (mean HbA1c change −0.4%, SD 1.0, P=.008). A mixed-effects model showed no significant treatment effect (IV vs UC) over 12 months (P=.06). However, participants with higher baseline HbA1c and those in the first time period experienced greater improvements in HbA1c. Both groups reported similar improvements in treatment satisfaction throughout the study.
Introduction

Poorly controlled diabetes, as indicated by elevated hemoglobin A1c (HbA1c), is associated with higher morbidity and mortality [1], greater cost of treatment [2], and poorer adherence to recommended self-management behaviors [3]. To improve HbA1c, diabetes self-management support needs to be accessible, responsive to varying patient health status, and effective in improving self-management skills, knowledge, and engagement. This is especially important for patients who struggle with self-management or face barriers to accessing traditional in-person services due to social determinants of health [4]. Integrated health care systems and payers, including commercial health plans, are particularly interested in innovative approaches to self-management support that address diabetes quality measures while reducing the overall cost of care [5]. Consequently, various commercial products have been developed to improve diabetes self-management, improve the experience of care, and reduce overall costs.

Electronic remote patient monitoring is a common strategy for many diabetes self-management applications available. This generally involves the transmission of self-monitored blood glucose readings to health care professionals and teams for evaluation and feedback [6]. Such real-time provider access to patient monitoring data presents an opportunity for care teams to deliver timely, tailored support without in-person contact. However, additional research targeting provider behavior with consideration of reimbursement for time and effort is needed to successfully integrate remote monitoring into routine care [7]. A recent meta-analysis of 4 systematic reviews of randomized controlled trials evaluating phone- and internet-based monitoring found improvement in HbA1c levels of −0.55% (95% CI −0.73 to −0.36) compared with usual care, though with statistical heterogeneity [6]. Notably, only 14 of 25 randomized trials reported significant improvement over usual care, with variability in what usual care support entails, as well as study quality. Potentially, positive findings may represent substandard care in comparison groups and may reflect the lack of resources required to ensure adequate evaluation and feedback is given to patients.

The Livongo for Diabetes Program is commercially available for purchase for individual use or can be implemented through a health organization or insurer. The program highlights the integration of Certified Diabetes Educators (CDEs), also referred to as Certified Diabetes Care and Education Specialists, who can provide real-time feedback on glucose monitoring data, including immediate responses to abnormal glucose excursions. One prior observational study of over 4500 individuals with diabetes using the Livongo for Diabetes Program found a decrease in glucose levels outside of a 70-180 mg/dL range [8]. However, the study did not include a comparison group to establish efficacy, and HbA1c was not assessed to understand if there was less hypoglycemia, less hyperglycemia, or both.

The present study was a randomized controlled crossover trial testing the efficacy of 6 months of participation in the Livongo for Diabetes Program in patients with poorly controlled type 2 diabetes. The primary outcome of the trial was change in HbA1c, with a secondary outcome of change in diabetes treatment satisfaction. In this study, we hypothesized that patients would experience greater improvements in HbA1c and treatment satisfaction when enrolled in the intervention program compared to usual care. Additionally, we explored engagement with the program, including monitor use and receipt of CDE support.

Methods

Setting and Recruitment

Participants with type 2 diabetes were recruited at the University of Massachusetts Medical Center Diabetes Center of Excellence (DCOE) from April 1 to July 9, 2015. All patients at the DCOE have both a primary care provider and a DCOE specialist provider. Inclusion criteria included the ability to speak English, a diagnosis of type 2 diabetes, and two consecutive HbA1c recordings greater than 8.0% in the previous 12 months, indicating poor glycemic control. Subjects were excluded if they were cognitively impaired (as designated by their provider), pregnant, or a prisoner. All human subjects research was reviewed and approved by the University of Massachusetts Medical School Institutional Review Board.

Research staff screened medical records of patients scheduled for routine appointments to identify those meeting the HbA1c criterion. The staff approached potentially eligible patients in the clinical environment and privately screened for eligibility if patients expressed interest. Patients were informed that they would be given access to the Livongo for Diabetes Program for a total of 6 months, either immediately or after a 6-month waiting period, randomly determined. Interested and eligible participants signed consent forms. Of 195 eligible subjects approached for recruitment, 123 (63.1%) expressed interest in participating, and 120 (61.5%) completed the informed consent process and were randomized to treatment groups. One subject failed to complete the baseline survey and was lost to follow-up prior to enrollment in the intervention.
Intervention

The intervention included free enrollment in the Livongo for Diabetes Program [9], the Livongo In Touch connected glucose meter, and a 6-month supply of testing supplies. The Livongo for Diabetes Program is accredited by the American Association of Diabetes Educators (AADE) Diabetes Education Accreditation Program and includes access to both scheduled and in-the-moment CDE support via phone call or SMS text messaging. At the time of the study, the Livongo for Diabetes program was not available as a direct-to-consumer product but was available to employees of several large companies.

The In Touch connected glucose meter is cellular-enabled, allowing for automatic uploading of self-monitoring of blood glucose (SMBG) recordings to a secure patient portal. Patients were instructed to use the meter to test their blood glucose as frequently as previously instructed by their providers. After patients use the meter to test their glucose, the SMBG recording is uploaded to the Livongo Smart Cloud. In this study, Livongo transferred all SMBG data to the DCOE electronic health record (EHR) system daily. The first time an uploaded blood glucose recording was above 250 mg/dL and anytime it was above 400 mg/dL or below 50 mg/dL, the Livongo Smart Cloud would notify the Livongo Care Team to perform outreach to the patient.

The Livongo Care Team of CDEs would contact participants by their preferred communication method (either phone call or text message) within 3 minutes of receiving an abnormal SMBG notification from the Smart Cloud. When contact was made, they would assess if the patient needed immediate medical attention, troubleshoot reasons for the flagged SMBG recording, and provide resources to improve self-management of diabetes. If a participant needed immediate medical attention, the CDE would direct them to call 911. If the intervention CDE believed a participant was in need of additional support from their DCOE care team, the CDE would contact the DCOE directly to request follow-up with the patient. Documentation of all encounters between intervention CDEs and participants was sent to the DCOE weekly to be entered into the EHR (Figure 1 for intervention components and flow of data). While the SMBG and CDE encounter data were available to the DCOE providers, the study did not target DCOE provider behavior (eg, by encouraging the providers to review or use the intervention data available in the EHR).

Intervention participants were encouraged at enrollment and during each CDE outreach to schedule follow-up coaching sessions with the CDEs. Coaching sessions covered the AADE’s 7 self-care behaviors: healthy eating, being active, glucose monitoring, taking medication, problem solving, reducing risks, and health coping [10]. While intervention CDEs did not give participants medical direction or make changes to their care plans, they answered diabetes-specific questions on topics such as nutrition and lifestyle changes and contacted the DCOE if they believed the participant would benefit from additional medical intervention.

Text-based messages sent to the participants through the meter after each test were based on the AADE National Standards for Diabetes Self-Management Education curriculum and included feedback and diabetes self-management tips. Other features of the meter included tagging SMBG recordings with contextual information (before meal, after meal, neither, and how they were feeling at the time of testing), an electronic logbook, and a built-in physical activity tracker. The meter also allowed participants to share SMBG data with their care providers or family via text message, email, or fax. While Livongo now

Figure 1. Intervention components and flow of patient data. CDE: Certified Diabetes Educator; SMBG: self-monitoring of blood glucose.
offers a mobile phone app to accompany the In Touch meter, this app was not available at the time of the study.

**Usual Care**

Participants in the usual care group continued to receive specialty care from DCOE and primary care providers. This included the recommended quarterly appointments with their DCOE care team and regular access to their providers through phone calls or secure messaging through the patient portal.

**Randomization**

A randomization table was created prior to the start of recruitment to equally allocate 120 participants to 2 treatment groups. The first group received the intervention for 6 months and then returned to usual care (IV/UC) for 6 months. The second group received usual care for 6 months before enrolling in the intervention (UC/IV) for 6 months. Study staff not involved with recruitment created enrollment folders for each participant based upon the randomization table. Study staff responsible for recruitment were blinded to treatment group designation from study enrollment during baseline questionnaire administration. For participants randomized to receive the intervention during the first time period, the last baseline survey item asked if they would like to schedule a phone call with research staff to walk through using the connected glucose meter when they received it at home. Those interested were scheduled for a tutorial approximately 7 days later, after confirmed delivery of the intervention start-up package containing the connected glucose meter and testing materials. A similar tutorial request process occurred at the end of the 6-month survey for participants receiving the intervention during the second time period.

**Data Collection**

At study enrollment, participants had an HbA1c test drawn. Participants were scheduled to return at 3, 6, 9, and 12 months ±1 week post–study enrollment for HbA1c testing. For participants who did not return for their scheduled 6-month (23/119, 19.3%) and 12-month (34/119, 28.6%) test, an HbA1c recording from their closest clinical visit was extracted from the EHR if it was within 3 months of the scheduled lab testing date (49/57, 86% of total missing). For patients without an available HbA1c in the EHR (8/57, 14% of total missing), change in HbA1c was imputed with the mean of their treatment group in mixed-effects modeling analyses.

Participants completed paper questionnaires at baseline, 6 months (prior to treatment crossover), and 12 months (study completion). Participants were administered questionnaires at the clinic and could finish them at home and mail them back, if necessary. Data from the questionnaires were manually entered by study staff using REDCap data capture tools [11]. Data on engagement with intervention, including number of SMBG recordings, number of CDE contacts, and number of CDE coaching sessions were collected by Livongo and securely transferred to study staff for manual entry into the REDCap project.

**Primary and Secondary Outcomes**

Changes in HbA1c during each time period were the primary outcomes of this study. HbA1c change was evaluated by comparing the mean changes in HbA1c while receiving the IV compared to HbA1c change while receiving UC. This was done for both the first treatment period and the second treatment period. Overall impact of the intervention on the change in HbA1c across both time periods was assessed in a mixed-effects model.

Diabetes treatment satisfaction was chosen as a secondary outcome because it is associated with positive diabetes outcomes, including HbA1c [12]. To measure baseline satisfaction with diabetes treatment, the Diabetes Treatment Satisfaction Questionnaire (DTSQ) was completed. The DTSQ is an 8-item measure with responses ranging from very satisfied to very dissatisfied for a total scale score range of 0 to 36 [13]. To evaluate change in satisfaction attributable to the intervention, the Diabetes Treatment Satisfaction Questionnaire Change (DTSQc) was included in the 6-month and 12-month questionnaires. The DTSQc is an 8-item measure that asks the extent to which participants experienced change in satisfaction over the course of the previous 6 months with responses ranging from much less satisfied now (−3) to much more satisfied now (+3) [14].

**Sample Size Estimation**

The primary outcome of this study was change in HbA1c. We anticipated the distribution of change in HbA1c would approximate a normal distribution, allowing for the use of a standard t test to examine differences in mean HbA1c change between treatment groups during each time period. Based on previous interventions in this patient population [15,16], we assumed a 1.0% difference in mean HbA1c change between treatment groups and a 1.5 SD in HbA1c change for both groups, requiring 48 participants per group for 90% power and a type I error rate of .05. We assumed a 10% dropout, which required 53 participants per arm. A conservative approach targeted recruitment of 60 participants per treatment group. Sample size calculations were performed using the SAMPSI command in Stata software, version 13.1 (StataCorp).

**Analytic Plan**

Bivariate comparisons of baseline characteristics between treatment groups were conducted to evaluate success of randomization. Baseline characteristics of the participants who failed to return for the 6-month and 12-month follow-up appointments were compared against those of participants who completed follow-up visits by using independent samples two-tailed t tests.

Primary outcome analyses involved independent samples two-tailed t tests to examine differences in HbA1c change between treatment groups during the first and second time periods. Both intent-to-treat and completer analyses were conducted. Participants were considered completers if they returned for the 6-month and 12-month follow-up visits. To account for the crossover design and multiple time points of the
study, a random intercept mixed-effects model with a restricted maximum likelihood estimator option of the mixed procedure in SAS software, version 9.4 (SAS Institute), was performed to examine variance between treatments by time with respect to subjects.

Results

Sample Characteristics
Study participants (n=119) had mean baseline HbA1c of 10.1% (SD 1.4). Age at enrollment ranged from 23 to 84 years old with an average age of 56.7 years (SD 11.6). The study sample was 52.9% (63/119) women and 71.4% (85/119) white (Table 1). Both groups were similar in terms of demographic characteristics, insulin use, HbA1c, and treatment satisfaction.
Table 1. Study participants’ characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>IV/UC&lt;sup&gt;a&lt;/sup&gt; (n=59)</th>
<th>UC/IV&lt;sup&gt;b&lt;/sup&gt; (n=60)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>56.1 (11.1)</td>
<td>57.4 (12.1)</td>
<td>.55</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>18-40</td>
<td>5 (8)</td>
<td>4 (7)</td>
<td>.35</td>
</tr>
<tr>
<td>40-65</td>
<td>42 (71)</td>
<td>39 (65)</td>
<td>.36</td>
</tr>
<tr>
<td>65+</td>
<td>12 (20)</td>
<td>17 (28)</td>
<td></td>
</tr>
<tr>
<td>Gender (women), n (%)</td>
<td>34 (58)</td>
<td>29 (48)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>.65</td>
</tr>
<tr>
<td>White</td>
<td>40 (68)</td>
<td>45 (75)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>6 (10)</td>
<td>3 (5)</td>
<td></td>
</tr>
<tr>
<td>Native/Alaskan American</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>More than 1 race</td>
<td>7 (12)</td>
<td>6 (10)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>5 (8)</td>
<td>6 (10)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td>.81</td>
</tr>
<tr>
<td>Hispanic Latinx</td>
<td>11 (19)</td>
<td>9 (15)</td>
<td></td>
</tr>
<tr>
<td>Not Hispanic Latinx</td>
<td>46 (78)</td>
<td>48 (80)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>2 (3)</td>
<td>3 (5)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td>.80</td>
</tr>
<tr>
<td>&lt;High school grad</td>
<td>9 (15)</td>
<td>7 (12)</td>
<td></td>
</tr>
<tr>
<td>High school grad</td>
<td>18 (31)</td>
<td>17 (28)</td>
<td></td>
</tr>
<tr>
<td>Post–high school trade</td>
<td>6 (10)</td>
<td>5 (8)</td>
<td></td>
</tr>
<tr>
<td>1-3 years college</td>
<td>14 (24)</td>
<td>16 (27)</td>
<td></td>
</tr>
<tr>
<td>College grad</td>
<td>11 (19)</td>
<td>13 (22)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td></td>
</tr>
<tr>
<td>Household income (US$), n (%)</td>
<td></td>
<td></td>
<td>.78</td>
</tr>
<tr>
<td>&lt;20k</td>
<td>24 (41)</td>
<td>22 (37)</td>
<td></td>
</tr>
<tr>
<td>20-50k</td>
<td>11 (19)</td>
<td>14 (23)</td>
<td></td>
</tr>
<tr>
<td>50-100k</td>
<td>10 (17)</td>
<td>11 (18)</td>
<td></td>
</tr>
<tr>
<td>&gt;100k</td>
<td>11 (19)</td>
<td>7 (12)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>3 (5)</td>
<td>6 (10)</td>
<td></td>
</tr>
<tr>
<td>Internet access, n (%)</td>
<td></td>
<td></td>
<td>.73</td>
</tr>
<tr>
<td>No</td>
<td>9 (15)</td>
<td>11 (18)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>50 (85)</td>
<td>47 (78)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>0 (0)</td>
<td>2 (3)</td>
<td></td>
</tr>
<tr>
<td>Insulin use, n (%)</td>
<td></td>
<td></td>
<td>.62</td>
</tr>
<tr>
<td>No</td>
<td>7 (12)</td>
<td>9 (15)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>52 (88)</td>
<td>51 (85)</td>
<td></td>
</tr>
<tr>
<td>HbA1c&lt;sup&gt;d&lt;/sup&gt;, %, mean (SD)</td>
<td>10.3 (1.4)</td>
<td>10.0 (1.4)</td>
<td>.21</td>
</tr>
<tr>
<td>Treatment satisfaction [14], mean (SD)</td>
<td>29.6 (5.3)</td>
<td>28.4 (5.2)</td>
<td>.24</td>
</tr>
</tbody>
</table>

<sup>a</sup>IV/UC: intervention for 6 months before usual care for 6 months.

<sup>b</sup>UC/IV: usual care for 6 months before intervention for 6 months.

<sup>c</sup>Not available.
**Study Retention**

Of the 119 study participants, 97 (81.5%) returned for the 6-month HbA1c lab, and 92 (77.3%) completed the 6-month follow-up survey (Figure 2). After treatment crossover, 86 (72.3%) participants returned for the 12-month HbA1c test, and 92 (77.3%) participants completed the 12-month follow-up survey. HbA1c data from the nearest clinical appointment were extracted for 19 of the 22 (86%) participants who did not return for the 6-month HbA1c lab and 30 of the 33 (91%) participants who did not return for the 12-month HbA1c lab. HbA1c values for the remaining participants with missing values at 6 months (n=3) and 12 months (n=3) were set to their group’s mean value so that the final analytic sample included follow-up HbA1c data for all 119 participants at the 6-month and 12-month time points.

**Figure 2.** Participant CONSORT (Consolidated Standards of Reporting Trials) diagram. HbA1c: hemoglobin A1c.

---

**Engagement With Intervention**

Among participants randomized to receive the intervention first (IV/UC, n=60), 1 (2%) did not enroll in the intervention program, and 6 (10%) never used the intervention meter. Of the 60 participants randomized to receive the intervention in the second period (UC/IV), 11 (18%) did not complete the 6-month follow-up visit and subsequently failed to enroll in the intervention. Of those participants who enrolled in the intervention in the second period (n=49), 8 (16%) never used the meter.

Among all participants who used the intervention meter during either time period (n=94), the average number of SMBG recordings per participant over the 6-month intervention period was 220 (SD 165, range: 2-817). For these participants, 73 (78%) were contacted by an intervention CDE at least once in response to a high or low SMBG recording outside of range. Over the course of the entire study, 400 support contacts were attempted by intervention CDEs, with 295 (73.8%) successful contacts, defined as reaching the patient (phone call) or receiving a reply (text message). Of these, 183 (62.0%) were by phone, and 112 (38.0%) were by SMS text messaging. Among the 73 participants contacted in response to a flagged SMBG, 11 (15%)
scheduled at least one follow-up coaching session with an intervention CDE. Among those who completed a coaching session with an intervention CDE, the average number of coaching sessions was 2.5 (SD 1.5) with a range from 1 to 5 total coaching sessions.

**Change in HbA1c**

Similar rates of HbA1c change were seen between both groups after 6 months ($t_{114}=1.06$, $P=.29$), with the intervention improving mean HbA1c by 1.1% (SD 1.5; $P<.001$) and usual care by 0.8% (SD 1.5; $P<.001$) (Table 2). After crossover, those returning to usual care (IV/UC) did not experience significant change in mean HbA1c ($P=.41$), while those who began receiving the intervention (UC/IV, n=39) had additional improvement in mean HbA1c by 0.4% (SD 1.0; $P=.008$) (Figure 3). The difference in mean HbA1c change during the second time period between groups was not statistically significant in intent-to-treat analyses ($P=.09$) but was significant among the participants who completed the final study visit ($P=.03$) (Table 2).

**Table 2.** Change in HbA1c percentage and diabetes treatment satisfaction, by group.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>IV/UC</th>
<th>UC/IV</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean (SD)</td>
<td>n</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HbA1c</td>
<td>59</td>
<td>10.3 (1.4)</td>
<td>60</td>
</tr>
<tr>
<td>DTSQ</td>
<td>56</td>
<td>29.6 (5.3)</td>
<td>59</td>
</tr>
<tr>
<td>6-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ HbA1c % from baseline (ITT)</td>
<td>56</td>
<td>−1.1 (1.5)</td>
<td>60</td>
</tr>
<tr>
<td>Δ HbA1c % from baseline (completer)</td>
<td>47</td>
<td>−1.1 (1.5)</td>
<td>49</td>
</tr>
<tr>
<td>DTSQc</td>
<td>42</td>
<td>+12.9 (5.5)</td>
<td>46</td>
</tr>
<tr>
<td>12-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ HbA1c % from 6-month (ITT)</td>
<td>56</td>
<td>+0.2 (1.7)</td>
<td>60</td>
</tr>
<tr>
<td>Δ HbA1c % from 6-month (completer)</td>
<td>41</td>
<td>+0.3 (1.7)</td>
<td>39</td>
</tr>
<tr>
<td>DTSQc</td>
<td>40</td>
<td>+11.5 (6.8)</td>
<td>42</td>
</tr>
</tbody>
</table>

$a$IV/UC: intervention for 6 months before usual care for 6 months.

$b$UC/IV: usual care for 6 months before intervention for 6 months.

$c$HbA1c: hemoglobin A1c.

dDTSQ: Diabetes Treatment Satisfaction Questionnaire.

eITT: intent-to-treat.

fDTSQc: Diabetes Treatment Satisfaction Questionnaire Change.
The mixed-effects model (Table 3) showed a nonsignificant difference in HbA_1c improvement of 0.4% between the intervention and usual care treatment conditions (P=.06). The model also showed significant effects of baseline HbA_1c (P=.03) and time period (P<.001). Participants with higher baseline HbA_1c saw greater HbA_1c improvement across the whole study, and there was greater HbA_1c improvement in the first period compared to the second period.

Table 3. Results of crossover (mixed-effects model) analysis of HbA_1c change.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HbA_1c(^a) % change estimate</th>
<th>SD</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline HbA_1c</td>
<td>−0.15</td>
<td>0.07</td>
<td>.03</td>
</tr>
<tr>
<td>Treatment (IV(^b) vs UC(^c))</td>
<td>−0.37</td>
<td>0.19</td>
<td>.06</td>
</tr>
<tr>
<td>Time period (1 vs 2)</td>
<td>−0.84</td>
<td>0.20</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Treatment × period</td>
<td>0.29</td>
<td>0.39</td>
<td>.46</td>
</tr>
</tbody>
</table>

\(^a\)HbA_1c: hemoglobin A_1c.

\(^b\)IV: intervention.

\(^c\)UC: usual care.

Change in Diabetes Treatment Satisfaction

Among participants completing the 6-month questionnaire (n=96), those receiving the intervention reported a mean improvement in treatment satisfaction of +12.9 (SD 5.6) compared to +10.7 (SD 6.6) with usual care (P=.09). Among those completing the final questionnaire (n=82), those who returned to usual care in the second time period (IV/UC) reported an improved mean treatment satisfaction change score of +11.5 (SD 6.8) compared to +13.4 (SD 4.5) among participants who received the intervention in the second time period (UC/IV, P=.15).

Discussion

Principal Results

In this 12-month randomized crossover trial, we found that patients enrolled in a diabetes remote monitoring program experienced improvements in HbA_1c and treatment satisfaction similar to usual care at a specialty diabetes center. Our mixed-effects model assessing HbA_1c change over both 6-month time periods estimated that HbA_1c improvement produced by the intervention was approximately 0.4% greater than that produced by usual care, though not reaching statistical significance (P=.06). At the same time, we did not observe...
differences in treatment satisfaction between the program and usual care. Together, these findings provide additional evidence regarding the expected outcomes of a commercial remote monitoring program, which may be useful for health organizations and insurers to consider in making decisions for patient self-management support.

In the first 6 months, patients experienced improvement in HbA1c, including those receiving usual care, who exhibited improvement in mean HbA1c by ~0.8%. This is a common finding in comparable trials involving patients with uncontrolled diabetes and may result from multiple factors. First, improvement through usual care could be due to the Hawthorne effect [17]. Participants received additional attention and engaged frequently with research staff, they were called and reminded to return quarterly for HbA1c testing, and they knew they would receive the anticipated commercial intervention after 6 months. Second, patients received specialized care through the DCOE endocrinologists and may represent more intensive blood glucose management than typically experienced through the primary care setting. This and potential “spillover” effects may have additionally narrowed differences observed between treatment conditions. Finally, “regression to the mean” may have contributed to improvements in all patients by recruiting only those with higher baseline HbA1c levels to the study.

Comparison With Prior Work

As in other studies, patients who missed follow-up visits for data collection had higher baseline HbA1c levels. For these individuals, it is not clear that commercial programs adequately address the barriers to complex diabetes self-management behaviors and social determinants of health, particularly with remote CDE support. Program CDEs may not develop the same relationships with patients as health care team members or recognize cultural, regional, or other psychosocial issues that may influence glycemia. Unfortunately, in many health care settings these patients still tend to have high no-show rates for appointments, worse diabetes-related health outcomes, lower rates of SMBG testing, and greater medication nonadherence [18-20]. Similar interventions involving SMBG and targeting patients with poorly controlled diabetes have demonstrated improvement in health outcomes for this increasingly prevalent and costly patient population [15,16,21-27]. Unique to this intervention was the in-the-moment, virtual support provided in response to abnormal SMBG levels uploaded automatically by connected glucose meters. By contacting patients immediately after their blood glucose tests high or low, CDEs could offer timely support when patients may need it most (eg, immediate hypoglycemia treatment). The CDE could also take advantage of “teachable moments” to provide diabetes education and self-management training when there is greater attention [28]. During these unplanned opportunities, patients can gain a better understanding of why their blood glucose is outside of range and learn how best to prevent it from happening again in the future.

While timely CDE outreach may be useful for some patients, it could also prompt stress in those who may not want to be contacted when SMBG levels are out of range. To address this concern, participants could adjust the SMBG levels that would trigger CDE contact; however, no participants requested to do this during the study. This may be secondary to following a “default” (status quo bias) [29] or may be due to a lack of technological knowledge on how to fully operate the meter. As a result, it remains possible that individuals will avoid self-testing if they suspect their levels are more extreme to avoid CDE involvement, especially if they exhibit more risk-seeking behavior [30]. If true, it suggests that for future implementation, this option should be emphasized upon initial training or reassessed over time.

Similarly, we found that only a small proportion of participants scheduled an individual coaching session with a program CDE. Routine scheduled coaching sessions for all participants may further enhance delivery of diabetes self-management education and training in this population. Additionally, CDEs could contact and counsel patients who have not recorded an SMBG level over an extended period. Besides the CDEs, the program could encourage greater involvement of a patient’s care team and support system, including informal caregivers such as family members. Providing caregivers with electronic access to a patient’s SMBG recordings and tools to assist in disease management may improve the quality of support they provide and reduce their own caregiver burden. We did not investigate the effects of this intervention on caregiver support and burden, but this should be considered in a future study.

Strengths

There were several strengths in this study. We collected both physiological (HbA1c) and patient-reported (diabetes treatment satisfaction) outcomes. Prior study of the program only included detection of glucose levels outside of range and excluded treatment satisfaction [8]. Additionally, the randomized controlled crossover study design allowed for both between- and within-group comparisons. This provided a more comprehensive evaluation by time period, treatment, and sequence of treatment received. Finally, we built an application programming interface to allow the transfer of SMBG and CDE/patient interactions from the Livongo cloud-based system to the clinic’s EHR. This allowed for the intervention data to be accessible to the patients’ care teams between clinic appointments.

Limitations

There are several limitations of this study to consider. The intervention time period was relatively short (6 months) for a group of patients with poorly controlled diabetes receiving care at a specialty diabetes center. The limited exposure to the intervention did not allow for evaluation of a sustained intervention effect. In addition, as only patients receiving the intervention had SMBG recordings regularly uploaded, we did not compare frequency of blood glucose testing during intervention compared to usual care. More research is needed with longer durations of intervention treatment, as most studies are 12 months or less [6], and in other patient populations, as this study only focused on patients with poorly controlled diabetes and did not collect data on duration of diabetes at time of enrollment. Second, data analyzed are from 2015 to 2016,
and the intervention program has made several adaptations since study completion. Livongo has partnered with several companies recently, including Dexcom and their continuous glucose monitoring (CGM) devices and Fitbit with their physical activity trackers. Furthermore, Livongo recently merged with Teladoc Health, a leading telemedicine provider. Further study of Livongo’s effect after incorporating CGM devices, wearable devices with more telehealth human coaching activities and advanced decision support, is needed. This is especially important considering a very limited number of participants in this study took advantage of a scheduled coaching session.

As well, while accessibility to virtual diabetes care support programs like Livongo has increased recently, many patients may continue to face barriers accessing or affording such support. These access to care challenges limit the generalizability of the study to only patients with access to such programs. Additionally, this study did not target provider behavior. SMBG data was uploaded to the EHR daily, but optimizing the use of these data by the usual care team was not part of the intervention. In regard to retention, several participants failed to return for their 6-month visit (28%), with those in the UC/IV group never receiving the intervention during the second study time period. Lastly, there may have been carryover of treatment effects for participants who received the intervention first (IV/UC), especially considering the absence of a washout period in the study design.

Conclusions
We found that patients with poorly controlled diabetes enrolled in the commercial remote diabetes monitoring program experienced improvements in HbA1c similar to when they received usual care at a specialized diabetes center. Improved treatment satisfaction was also reported by both groups throughout the study. Further development targeting patient engagement in the program and access to CDEs for diabetes support could result in greater program impact, especially for patients with limited access to specialized diabetes care. Future interventions involving diabetes care monitoring programs and connected technologies should consider including a structured coaching component, proactively involving caregivers and family members of patients, and investing in additional efforts to engage patients who are more likely to miss scheduled study activities and appointments. Better integration of diabetes remote monitoring programs into routine clinical care must be prioritized. This is necessary in order to achieve the full potential benefit from similar interventions in the future. In addition, cost-effectiveness needs to be investigated. This will be critical in justifying the expense required to provide in-the-moment support offered by the intervention.

Acknowledgments
This study was jointly funded by Livongo Health and the UMass Medical School Diabetes Center of Excellence. This paper was supported by effort from grant KL2 TR001455. Livongo provides consent for the use of Figure 1 in this publication and reserves all rights to the Figure 1 and this consent shall not be deemed as Livongo providing any consent or future use of Figure 1 to the publication.

Authors’ Contributions
All authors have participated in the development of the intervention, analysis of results, or scientific writing of the paper.

Conflicts of Interest
DDM receives grant or consulting support from Bristol Myers Squibb, Boehringer Ingelheim, Pfizer, Philips, Samsung, Avania, Apple, Heart Rhythm Society, Fitbit, and Flexcon. DDM serves on the GUARD AF and Fitbit Heart Study Steering or Advisory Committees. BSG’s spouse is employed by Abbott Labs, which manufactures continuous glucose monitors.

Multimedia Appendix 1
CONSORT EHEALTH checklist.
[PDF File (Adobe PDF File), 12954 KB - diabetes_v6i1e25574_app1.pdf ]

References


Abbreviations

- AADE: American Association of Diabetes Educators
- CDE: Certified Diabetes Educator
- CGM: continuous glucose monitoring
- DCOE: Diabetes Center of Excellence
- DTSQ: Diabetes Treatment Satisfaction Questionnaire
- DTSQc: Diabetes Treatment Satisfaction Questionnaire Change
- EHR: electronic health record
- HbA1c: hemoglobin A1c
- IV: intervention
- UC: usual care
- SMBG: self-monitoring of blood glucose
- T2D: type 2 diabetes

©Daniel J Amante, David M Harlan, Stephenie C Lemon, David D McManus, Oladapo O Olaitan, Sherry L Pagoto, Ben S Gerber, Michael J Thompson. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 11.03.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Telemetric Interventions Offer New Opportunities for Managing Type 1 Diabetes Mellitus: Systematic Meta-review

Claudia Eberle¹, MD, Prof Dr; Stefanie Stichling¹, MSc
Medicine with Specialization in Internal Medicine and General Medicine, Hochschule Fulda - University of Applied Sciences, Fulda, Germany

Corresponding Author: Claudia Eberle, MD, Prof Dr
Medicine with Specialization in Internal Medicine and General Medicine
Hochschule Fulda - University of Applied Sciences
Leipziger Strasse 123
Fulda, 36037
Germany
Phone: 49 661 9640 ext 6328
Email: claudia.eberle@hs-fulda.de

Abstract

Background: The prevalence of diabetes mellitus (DM) is increasing rapidly worldwide. Simultaneously, technological advances are offering new opportunities for better management of type 1 diabetes mellitus (T1DM). Telemetry, the remote acquisition of patient data via a telecommunication system, is a promising field of application in eHealth and is rapidly gaining importance.

Objective: The aim of this study was to summarize the current evidences available on the effectiveness of telemetric approaches in T1DM management. This systematic meta-review examined different types of interventions of the technologies used in communication between health care professionals and patients as well as the key outcomes.

Methods: We performed a systematic search in Web of Science Core Collection, EMBASE, Cochrane Library, MEDLINE via PubMed, and CINAHL databases in April 2020 with regard to the effectiveness of telemetric interventions for T1DM. We classified the interventions into 4 categories according to the technology used: (1) real-time video communication, (2) real-time audio communication, (3) asynchronous communication, and (4) combined forms of communication (real-time and asynchronous). We considered various study designs such as systematic reviews, clinical trials, meta-analyses, and randomized controlled trials and focused on the key outcomes. Additionally, a funnel plot based on hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) values and different quality assessments were performed.

Results: We identified 17 (6 high quality and 9 moderate quality) eligible publications: randomized controlled trials (n=9), systematic reviews and meta-analyses (n=5), cohort studies (n=2), and qualitative publications (n=1). Of 12 studies, 8 (67%) indicated a (significant or nonsignificant) reduction in HbA\textsubscript{1c} levels; 65% (11/17) of the studies reported overall (mildly) positive effects of telemetric interventions by addressing all the measured outcomes. Asynchronous interventions were the most successful for patients diagnosed with T1DM, but no technology was clearly superior. However, there were many nonsignificant results and not sustained effects, and in some studies, the control group benefited from telemetric support or increased frequency of contacts.

Conclusions: Based on the currently available literature, this systematic meta-review shows that telemetric interventions cause significant reduction in HbA\textsubscript{1c} levels and result in overall positive effects in T1DM management. However, more specified effects of telemetric approaches in T1DM management should be analyzed in detail in larger cohorts.

(JMIR Diabetes 2021;6(1):e20270) doi:10.2196/20270

KEYWORDS

type 1 diabetes; telemetry; telemedicine; telemonitoring; digital health; eHealth; diabetes management; systematic meta-review

Introduction

The historical origins of digital health date back to the 1970s, when telematics, the science of telecommunications and informatics, emerged [1]. Telemedicine developed as a technology-supported physician-patient relationship in the 1970s/80s as a subarea of telematics. In the 1990s, the emergence of the internet resulted in new communication channels and the development of eHealth [1]. Mobile health, which was developed as a subarea of eHealth in 2010, is referred
Telemedicine is a digital field of application and part of eHealth and digitalization in the health care sector [4]. The exchange between different user groups (eg, physician, patient, service provider) takes place in these apps [5]. When integrating users in the area of eHealth, and thus in telemedicine, a distinction is made between different forms of communication structures. This review focuses on the communication structure of “physician to patient,” which defines the communication between physicians (or health care professionals) and patients [5]. Telemetry has the advantage that no physical presence is necessary [6]. Telemetry is characterized by the American Telemedicine Association as “remote acquisition, recording, and transmission of patient data via a telecommunications system to a health care professional for analysis and decision making” [6]. In telemetric interventions, patients upload data (eg, dietary habits and glucose levels) and health care professionals review these data and offer feedback (eg, regarding medication and lifestyle) [6,7]. In this regard, telementoring describes the use of telecommunications (eg, audio or video) and electronic information processing technologies to provide those customized instructions [6].

This systematic meta-review focuses on telemetry by using the example of patients diagnosed with type 1 diabetes mellitus (T1DM). DM is one of the most prevalent chronic diseases worldwide [8]. Globally, approximately 463 million adults (age range 20-79 years) are diagnosed with DM [8]. T1DM accounts for 5%-10% of all DM forms and can arise at any age; however, it is frequently reported in kids and young adults [8]. The prevalence of T1DM has been increasing in the past decades. Globally, about 1.1 million children and adolescents (age range 0-19 years) are diagnosed with T1DM [8]. From a pathophysiological and a clinical view, T1DM is a very complex disease, which is dependent on beta-cell demolition by the T cells of the immune system, resulting in the total lack of insulin [9]. Comorbidities such as microvascular (eg, nephropathy, retinopathy, and neuropathy) and macrovascular (eg, cardiovascular disease, stroke) complications are closely and frequently related to DM [9]. Optimal glycemic control is the therapy goal to reduce and prevent such diabetic complications and comorbidities. Intensive therapeutic measures address the delay of onset of diabetic complications as well as comorbidities in T1DM [10]. Therefore, technological advances in diabetes therapy may provide powerful novel solutions for a better and more closed-meshed disease management [11]. Several studies have examined the capability of telemetry in the treatment of DM [12-14]. The use of technological apps may be an attractive option for T1DM management. Previous studies have shown feasibility and satisfaction by using telemedicine [13,14]. However, the evidence for the impact of telemetric interventions in the context of diabetes therapy and the potential of these interventions should be examined further. Therefore, this systematic meta-review intended to assess the current evidence for the effectiveness of telemetric interventions in the management of T1DM. Not only randomized controlled trials (RCTs), as it is often the case in the literature, but also various study designs, including clinical trials, systematic reviews, and meta-analyses, were considered.

### Methods

#### Search Strategy

We performed a systematic search in Web of Science Core Collection, EMBASE, Cochrane Library, MEDLINE via PubMed, and CINAHL databases in April 2020. The systematic meta-review was carried out based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [15]. Peer-reviewed full-text publications assessing the effectiveness of telemetric interventions in patients with T1DM, published from 2008 to April 2020, were included. We selected keywords from the medical subject headings and EMBASE subject headings databases and used title/abstract terms. The following Boolean logic was applied: (Diabetes Mellitus) AND (Telemetry OR Telemonitoring OR Telemedicine). No restrictions for geographical locations were placed. Initially, we carried out an extensive literature search with a strategy that covered different types of DM (T1DM, type 2 DM [T2DM], and gestational DM). During the process, T1DM studies were selected for this systematic meta-review. We additionally carried out manual researches of the references of the included examinations to recognize other reasonable publications. All search terms for the individual databases are provided in Multimedia Appendix 1.

#### Inclusion Criteria

We included publications written in English and German with the target group patients diagnosed with T1DM. These publications addressed interventions in the field of telemetry, telemedicine, and telemonitoring for their diabetes therapy. The intervention involved direct interaction between the patients and health care professionals, that is, feedback from health care professionals based on the transmitted patient data. We included the following study designs: systematic reviews, meta-analyses, clinical trials, and RCTs.

#### Exclusion Criteria

Since this systematic meta-review focused on T1DM, we excluded participants diagnosed with other forms of DM (such as T2DM, gestational DM, and other types of diabetes) as well as mixed collectives, meaning that studies included not only patients with T1DM but also people diagnosed with other types of DM. Moreover, we excluded individual studies that were already included in the identified systematic reviews and meta-analyses; therefore, no data from systematic reviews/meta-analyses and individual studies are pooled, leading to a possible bias. Abstracts, posters, comments, letters, study protocols, notes, and proceedings papers were excluded. In addition, publications that focused on the description of the technology were rejected. Telemetry is a wide term and may cover different technologies. Since the way of communication between patients and health care professionals is different compared to that in telemetric interventions, we analyzed...
interventions with mobile apps in other studies separately. We also eliminated studies providing only pooled data (ie, with patients of other diseases and with digital apps other than telemetry). Furthermore, duplicates and studies that addressed prevention or diagnosis of DM were rejected. The literature search is documented in the PRISMA flowchart (Figure 1). As Figure 1 shows, we selected T1DM studies from our extensive literature search.

Figure 1. PRISMA flowchart of the procedure for the search and selection of suitable publications (adapted from Moher et al [15]). GDM: gestational diabetes mellitus; T1DM: type 1 diabetes mellitus, T2DM: type 2 diabetes mellitus.

Data Extraction
We extracted the year of publication, study designs, durations, intervention and control groups, outcome measures, sample sizes, country, statistical significances, and conclusions. Intervention and control group data included the technologies used, feedback methods, the frequency of contact, and data transmission. The significance involved the comparison of the intervention group with the control group (intergroup) and the comparison within the intervention group, that is, from the baseline to the end of the study (intragroup), depending on what was reported. In relation to the systematic reviews and meta-analyses, the overall effects were extracted (overall positive effect, no effect, or inconclusive results). The quality of life (QoL) was divided into diabetes-related quality of life (DRQoL) as well as health-related quality of life (HRQoL).

Data Synthesis and Analysis
A qualitative analysis was conducted. The selected studies differed regarding sample, design, and measures. A proper meta-analysis was therefore not possible. For analysis, the studies were classified into different categories based on a
scheme that we developed. First, the publications were systematized into 4 categories according to the technologies used to communicate between the health care professionals and the patients (Textbox 1).

**Textbox 1.** Categories for the classification of the different intervention types.

<table>
<thead>
<tr>
<th>Different intervention types</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Real-time communication video: Synchronous face-to-face communication by videoconferencing and videoconsulting.</td>
</tr>
<tr>
<td>• Real-time communication audio: Synchronous communication by telephone calls (telephone coaching and counselling).</td>
</tr>
<tr>
<td>• Asynchronous communication: Asynchronous communication by email, SMS text messaging, internet/web-based platforms, server, home gateway, or post.</td>
</tr>
<tr>
<td>• Combined forms of communication: The intervention involves real-time and asynchronous communication.</td>
</tr>
</tbody>
</table>

Due to the heterogeneity, systematic reviews and meta-analyses were not assigned to these categories. Second, the studies were differentiated according to their designs. Third, these were structured based on key outcomes: hemoglobin A\(_1c\) (HbA\(_1c\)), body weight, blood pressure, QoL, cost-effectiveness, and time saved (Figure 2).

**Figure 2.** Scheme for structuring the included studies. BP: blood pressure; HbA\(_1c\): hemoglobin A\(_1c\); DRQoL: diabetes-related quality of life; HRQoL: health-related quality of life; MA: meta-analysis; SR: systematic review; RCT: randomized controlled trial.

---

**Assessment of Risk of Bias**

A quality assessment of the studies was conducted to determine the risk of bias. Since we included different study designs, we applied 3 different quality appraisal tools. First, we applied A MeaSurement Tool to Assess systematic Reviews (AMSTAR 2), a validated and widely used tool for the evaluation of systematic reviews and meta-analyses. AMSTAR 2 rates the study quality as high, moderate, low, or critically low. Second, we used Effective Public Health Practice Project (EPHPP), a validated instrument that addresses studies on health-related topics. Since this tool is suitable for quantitative intervention studies, we used it for RCTs and cohort studies. EPHPP consists of the following components: selection bias, study design, confounders, blinding, data collection methods, and withdrawals and drop-outs. The instrument rates the study quality as strong, moderate, or weak. Third, we applied the validated National Institute for Health and Care Excellence (NICE) quality appraisal checklist for qualitative studies. The NICE checklist includes the following components: theoretical approach, study design, data collection, trustworthiness, analysis, and ethics. This tool rates the study quality as ++ (high), + (moderate), or − (low). In addition, the publication bias was assessed visually as a funnel plot by using HbA\(_1c\) values. The studies were extremely heterogeneous. Without systematic reviews, meta-analyses, and cohort studies (ie, without control group) and excluding a study that compared 2 telemetric applications, we generated a funnel plot based on 6 RCTs. Intervention effect was expressed as the mean difference using HbA\(_1c\) values at the end of the study.
Results

Study Characteristics

The database search resulted in 1647 records. After removing duplicates, 1116 publications were screened for eligibility. We excluded 875 of these records based on titles/abstracts for the reasons given in Figure 1. After reviewing 241 full-text publications and an additional research of reference lists, a total of 189 studies were identified (T1DM, n=23; T2DM, n=99; gestational DM, n=11; and both T1DM/T2DM, n=51). We excluded 6 individual studies [16-21] that were already involved in the systematic reviews/meta-analyses. Finally, 17 publications were included in this synthesis. Multimedia Appendix 2 provides a detailed summary of each publication selected for inclusion in this systematic meta-review, including all measured outcomes. Table 1 shows the features of the included studies. Most studies (with exception of systematic reviews and meta-analyses due to their heterogeneity) were performed in Europe (n=6), followed by in the United States (n=3), Asia (n=1), and Russia (n=1), along with not specified (n=1). We categorized the studies by the type of intervention: real-time communication via video (n=3), asynchronous communication (n=4), and combined forms of communication (n=4). One qualitative study did not explain the intervention in detail. No real-time audio interventions were identified. Most studies were RCTs (n=9), systematic reviews and meta-analyses (n=5), as well as cohort studies (n=2), and qualitative publications (n=1). A presentation of all the intervention effects (significant and nonsignificant) on the key outcomes is provided in Multimedia Appendix 3. Two systematic reviews and meta-analyses were assessed as high-quality studies, whereas 2 were rated as moderate and 1 as critically low quality. Of the real-time video interventions, 3 were high-quality studies. Furthermore, 4 asynchronous interventions were rated as moderate quality. Of the combined interventions, 1 was rated as high, 2 as moderate, and 1 as weak-quality study. In addition, the qualitative publication was of moderate quality. The detailed quality appraisals are presented in Multimedia Appendix 4.

Table 1. Baseline characteristics of all the included publications.

<table>
<thead>
<tr>
<th>Characteristics of the publications</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study design (n=17)</strong></td>
<td></td>
</tr>
<tr>
<td>Systematic reviews and meta-analyses (total)</td>
<td>5 (29)</td>
</tr>
<tr>
<td>Randomized controlled trial (total)</td>
<td>9 (53)</td>
</tr>
<tr>
<td>Cohort (total)</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Qualitative (total)</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Year of publication (n=17)</strong></td>
<td></td>
</tr>
<tr>
<td>2008-2011</td>
<td>2 (12)</td>
</tr>
<tr>
<td>2012-2014</td>
<td>4 (24)</td>
</tr>
<tr>
<td>2015-2017</td>
<td>5 (29)</td>
</tr>
<tr>
<td>2018-2020</td>
<td>6 (35)</td>
</tr>
<tr>
<td><strong>Excluding systematic reviews and meta-analyses (n=12)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Europe</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Asia</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Russia</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Intervention type</strong></td>
<td></td>
</tr>
<tr>
<td>Real-time video</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Asynchronous</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Combined forms</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

a This included 1 pilot randomized controlled trial.
b This included 1 pilot cohort study.

HRQoL and DRQoL were evaluated using very different methods. Validated instruments were used to measure these outcomes, for example, 36-item Short Form Health Survey, Diabetes Quality of Life questionnaire, PedsQLTM 3.0 Diabetes Module questionnaire, 12-item Short Form Health Survey, and
European Quality of Life survey. There were also specially designed questionnaires.

**Effectiveness of Telemetry: Key Outcomes**

Of 17 studies, 11 (65%) reported overall (mildly) positive effects of the telemetric interventions in relation to all measured outcomes (Multimedia Appendix 2). Table 2 presents the significant effects (intragroup and intergroup) on the key outcomes. Of 12 studies, 8 (67%) indicated a (significant or nonsignificant) reduction (intragroup or intergroup) in HbA$_{1c}$ levels in the intervention group. Descriptive examination of the funnel plot by using HbA$_{1c}$ values based on 6 RCTs indicated a mild form of asymmetry (Multimedia Appendix 5).

**Table 2.** Impact of the interventions on selected outcomes (intragroup and intergroup) (n=17).$^a$

<table>
<thead>
<tr>
<th>Outcomes/ interventions</th>
<th>Hemoglobin A$_{1c}$</th>
<th>Blood pressure</th>
<th>Body weight</th>
<th>Diabetes-related quality of life</th>
<th>Health-related quality of life</th>
<th>Costs</th>
<th>Time saved</th>
<th>Others or not significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic review and meta-analysis</td>
<td>3</td>
<td><em>b</em></td>
<td>_</td>
<td>1</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Real-time video$^c$</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Asynchronous</td>
<td>1</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Combined</td>
<td>1</td>
<td>_</td>
<td>_</td>
<td>1</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Not specified$^c$</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
</tbody>
</table>

$a$All studies that reported significant intervention effects are mentioned in this table, including those effects that were not sustainable. This table does not include studies reporting nonsignificant intervention effects. The values in the tables indicate the number of studies that examined the outcome and these studies showed improvement in that particular outcome.

$b$Not available.

$c$Studies in this category did not examine any of the listed outcomes nor report any significant effects.

**Systematic Reviews and Meta-Analyses**

**HbA$_{1c}$ Levels (n=5)**

All 5 systematic reviews and meta-analyses analyzed HbA$_{1c}$ levels as the targeted outcome. Three studies (60%) reported overall positive effects in terms of reducing HbA$_{1c}$ levels significantly. Lee et al [12] (high-quality study) described a mean reduction of 0.18% (95% CI 0.04-0.33, $P$=.01). Peterson [22] (critically low-quality study) outlined that 12 studies showed a decline in HbA$_{1c}$ levels in their intervention groups. However, Viana et al [23] (moderate-quality) and Shulman et al [24] (high-quality study) found no significant decrease in HbA$_{1c}$ levels following telemedical interventions (mean deviation –0.124%, 95% CI, –0.268 to 0.020; $P$=.09 [25] and mean deviation –0.12, 95% CI, –0.35 to 0.11; $P$>.05 [24], respectively).

**Blood Pressure and Body Weight (n=1)**

Lee et al [12] (high-quality study) observed no benefits through telemedicine on either blood pressure or body weight.

**DRQoL (n=3) and HRQoL (n=1)**

Three studies examined the DRQoL. Two high-quality studies (67%) found no effects [12,24] and a moderate-quality review [26] that only included 1 suitable study found a significant improvement in DRQoL. In addition, 1 review observed no benefits on generic HRQoL [12].

**Cost-Effectiveness (n=1)**

One high-quality study described that the limited data available on the costs of telemedicine suggested no differences between the groups [24]. One of the included studies of this review reported that the intervention group omitted the 3-month visit, which saved US $142 [24].

**Asynchronous Interventions**

**HbA$_{1c}$ Levels (n=3)**

A cohort study (moderate quality) reported significantly reduced mean HbA$_{1c}$ levels at the end of the assessment phase ($P$=.01) [27]. However, another 2 moderate-quality RCTs found no significant differences HbA$_{1c}$ values between groups ($P$=.84 [28] and $P$=.49 [29]). One of these studies [28] examined telemedicine in addition to conventional care in the intervention group.

**HRQoL (n=1)**

One moderate-quality RCT observed that changes in HRQoL between the first visit and the final visit did not differ between the groups [30].

**Combined Interventions**

**HbA$_{1c}$ Levels (n=4)**

All 4 RCTs considered the outcome HbA$_{1c}$. Only 1 study (moderate quality) showed significant improvements in the HbA$_{1c}$ levels in the patients undergoing interventions (8.7% to 7.7%) compared to the controls (8.7% to 8.4%, $P<.05$) [31]. Gandrud et al (weak-quality study) [32] and Yaron et al [25] (high-quality study) reported positive but no significant differences in the effects on HbA$_{1c}$ levels between the telemedicine and usual care groups. In addition, 1 moderate-quality publication mentioned no improvement in HbA$_{1c}$ levels, with no statistically significant difference ($P$=.56
for control group, $P=.45$ for telemetry group, and $P=.60$ between groups) [33].

**DRQoL (n=2)**

According to an RCT (weak-quality study), a number of QoL indicators increased significantly due to telemetry compared to that in the control group ($P<0.05$) [31]. However, another moderate-quality RCT showed no significant increase in QoL by 6.5 points and 1.3 points for intervention group and control group ($P=0.06$, respectively [32].

**Cost-Effectiveness (n=2) and Time Saved (n=1)**

Yaron et al (high-quality study) [25] and Bertuzzi et al (moderate-quality study) [33] reported a cost reduction through telemedicine (no significance reported). Direct expenses were 24% lesser in the intervention group, while indirect costs diminished by 22% [25]. One of these studies also mentioned that patients saved time for each visit (mean 115 [SD 86] min) [33].

**Discussion**

**Principal Results**

This systematic meta-review highlighted the variety of telemetric interventions and technologies used in diabetes care by focusing on T1DM management. Considering all the study designs, asynchronous interventions were found to be the most successful for people with T1DM in improving the key diabetic outcomes, but no technology was clearly superior. However, the results might be inconsistent in terms of the different key outcomes, but fortunately, an improvement in terms of HbA1c values was found. HbA1c was by far the most investigated outcome in these studies. Overall, most systematic reviews and meta-analyses (high and moderate quality) showed a significant reduction in HbA1c values. The other systematic reviews and meta-analyses also indicated positive effects, but they were not statistically significant. The study of Lee et al [12], a high-quality study, achieved a significant and clear reduction of $-0.18\%$ (95% CI 0.04-0.33, $P=0.01$). Moreover, HbA1c levels were improved significantly in most asynchronous interventions. HbA1c values clearly decreased when combined interventions (asynchronous and real-time communication) were applied, but 1 moderate-quality study showed significant improvements and 3 more (high, moderate, and weak quality) reported positive but not significant effects. Our findings indicated a trend toward better glycemic control for patients with T1DM by means of telemedicine. This result has potential practical implications. The fact that HbA1c levels could be significantly improved in many studies is a promising result in view of the fact that an optimized glycemic control reduces the risk of comorbidities and complications as well as progression of microvascular and macrovascular consequences among patients with T1DM [10]. However, there are only few results for the other outcomes to be able to reach firm inferences. Blood pressure and body weights were examined by 1 meta-analysis. Lee et al (high-quality study) noticed that there are only few studies available revealing no obvious benefits [12]. Aside from that, 2 systematic reviews and meta-analyses (high and moderate quality) outlined no effects in terms of QoL, but a moderate-quality study demonstrated positive tendencies in improving the QoL. Overall, the studies reported that data availability is limited and further investigations are needed. Besides, DRQoL improved significantly in the “real-time video intervention” with weak quality. The moderate-quality asynchronous intervention showed no differences in HRQoL. However, DRQoL also improved obviously in combined interventions, that is, significantly in a weak-quality study and not significantly in a moderate-quality study. In general, there were only few studies on the cost-effectiveness of telemetric interventions. Costs were significantly reduced through “asynchronous interventions,” which was shown by a high-quality study. This high-quality study also demonstrated significant time saving through the asynchronous intervention. With combined interventions, 2 moderate-quality studies also showed clear cost reductions.

In our view, telemetry enables close diabetes management and offers the advantage of overcoming the physical presence. Telemetric technologies allow a higher frequency of contacts between patients and health care professionals. Telemetric interventions also increase, in our view, patient compliance, reliance, and empowerment. The patients implement recommendations for action more successfully in everyday life. They are supervised and managed effectively and more closely and may feel more secure in terms of diabetes therapy. Another systematic review and meta-analysis [12] that recently examined telemetry for the management of clinical outcomes of T1DM also showed that the evidence regarding body weight and blood pressure is clearly limited. In practice, considering the restricted availability of resources, it is important whether the telemetric interventions are cost-effective and time-saving. Therefore, these outcomes are of major importance and should be considered more often in studies in future. Interestingly and surprisingly, fasting blood glucose values seem to be a neglected outcome in these T1DM studies. Since accurate blood sugar measurements are required to reach euglycemic conditions with appropriate insulin doses [9], this outcome is very important.

The systematic reviews and meta-analyses were heterogeneous since telemetry can cover various interventions and technologies and the authors used different definitions of telemedical approaches. Additionally, the variability of the methods used in the studies made it difficult to reach firm conclusions. Studies often suffered from small sample sizes, poor study designs, lack of controls, or no long-term intervention effects. Some studies had samples of patients with poorly controlled diabetes that led to greater intervention effects. Overall, there were not many significant results both for intergroup and intragroup comparisons.

Interestingly, the control group was often not a real or pure control group with usual care. The control group often had an increased frequency of contacts with health care professionals (more than 4 times a year), which led to improved outcomes. In some studies, the control group benefited from telemetric support. Moreover, several studies did not adequately define usual care. The intervention effects might be greater if the telemetric group was compared to a pure control group. Besides, the high number of nonsignificant results is particularly
noticeable. This could be related to an often low statistical power. It is also concerning that some studies did not publish $P$ values. Furthermore, based on the findings, the long-term effects can be questioned. Some studies found significant positive postintervention effects, but they did not last for a long term. Long follow-up periods are therefore important.

Our review is, as far as we know, the first systematic meta-review on telemedicine in T1DM management. Compared to other papers, this systematic meta-review included different study designs, looked at a variety of outcomes, and carried out a differentiated analysis based on a developed scheme. We also analyzed the findings in detail and differentiated them based on the intergroup or intragroup comparison, significant or not significant effects, and effect sizes. In this way, we were able to contribute to a multifaceted view of the topic.

Limitations
Some limitations have to be considered when interpreting and using the results. To the best of our knowledge and the elected inclusion and exclusion criteria, we included all suitable studies. Some of the systematic reviews and meta-analyses reported that the poor quality of the included studies was a weakness. Furthermore, numerous definitions of telemetry and telemedicine include different technologies. For the reasons mentioned above, we decided to exclude smartphone app–based interventions, which may be a limitation. Besides, the definition of usual care was insufficient and heterogeneous across the publications. Some studies did not use a control group in the sense of usual care. It is notable that in some studies, the control group had a similar frequency of contacts as the intervention group. In some studies, the control group received telemetric support. These circumstances influence the results achieved and must be considered. Overall, the studies displayed different characteristics and methods, which lead to heterogeneity and can influence the reliability of the results.

Comparison With Prior Work
In a nutshell, other reviews showed similar inconsistent findings. Lee et al [12] observed no benefits in the interventions with telemedicine focused on blood pressure, body weight, and QoL in 38 RCTs. The overall value of the included interventions was insufficient for glycemic control and other clinical outcomes among patients with T1DM. Viana et al [23] examined telecare interventions to improve patients’ compliance and HbA$_1c$ values and found no decrease in HbA$_1c$ levels after telecare ($P=.09$).

Another systematic review [34] mentioned that 7 of the 14 included publications indicated statistically significant decreases in the observed outcomes, while 79% mentioned success with their telemetric interventions. Baron et al [35] investigated the effectiveness of mobile monitoring technologies for HbA$_1c$ levels in 24 studies and found inconsistent evidence for T1DM.

Conclusions
This systematic meta-review offered a comprehensive summary of the effectiveness of telemetric interventions in T1DM management and provided insights into the application of telemetric interventions. The evidence for the effectiveness of telemetric approaches in the management of T1DM might be inconsistent. Further studies with a clear and homogeneous methodology are necessary for research and for patients. In addition, we need further research to understand how, why, and when technology can improve the outcomes. Studies should not only focus on HbA$_1c$ but also address other outcomes, in particular, fasting blood glucose, blood pressure, QoL, cost-effectiveness, and time saved. Additionally, future studies should provide sufficient statistical power. Further research regarding T1DM is required to examine the special needs of this subgroup in more detail and to develop and adapt suitable interventions. The alarming number of findings with nonsignificant $P$ values reveals a need for better study planning as well as RCTs with large sample sizes. In conclusion, telemetry might be a promising approach for people diagnosed with T1DM, especially asynchronous interventions, but its potential should be explored further.

Acknowledgments
This manuscript was created in the context of the project with the number EB 440/4-1 by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). Therefore, we would like to thank the DFG for strongly supporting this research work.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search terms for the databases.
[PDF File (Adobe PDF File), 91 KB - diabetes_v6i1e20270_app1.pdf ]

Multimedia Appendix 2
Detailed summary of each publication selected for inclusion in the systematic meta-review, including all measured outcomes ($n=17$).
[PDF File (Adobe PDF File), 606 KB - diabetes_v6i1e20270_app2.pdf ]
Multimedia Appendix 3
Detailed presentation of all intervention effects (significant and nonsignificant) on the key outcomes.
[PDF File (Adobe PDF File), 446 KB - diabetes_v6i1e20270_app3.pdf ]

Multimedia Appendix 4
Quality assessment using A MeaSurement Tool to Assess systematic Reviews (AMSTAR 2) (n=5 studies).
[PDF File (Adobe PDF File), 481 KB - diabetes_v6i1e20270_app4.pdf ]

Multimedia Appendix 5
Funnel plot assessing publication bias using HbA1c levels (%) at the end of the study.
[PDF File (Adobe PDF File), 621 KB - diabetes_v6i1e20270_app5.pdf ]

References


**NICE:** National Institute for Health and Care Excellence

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**QoL:** quality of life

**RCT:** randomized controlled trial

**T1DM:** type 1 diabetes mellitus

**T2DM:** type 2 diabetes mellitus

Edited by C Richardson; submitted 19.05.20; peer-reviewed by E van der Velde, E Burner; comments to author 29.06.20; revised version received 20.07.20; accepted 16.02.21; published 16.03.21.

Please cite as:
Eberle C, Stichling S
Telemetric Interventions Offer New Opportunities for Managing Type 1 Diabetes Mellitus: Systematic Meta-review
JMIR Diabetes 2021;6(1):e20270
URL: https://diabetes.jmir.org/2021/1/e20270
doi:10.2196/20270
PMID:33724201

©Claudia Eberle, Stefanie Stichling. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 16.03.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Exchanges in a Virtual Environment for Diabetes Self-Management Education and Support: Social Network Analysis

Carlos A Pérez-Aldana¹, MSc; Allison A Lewinski²,³, PhD, MPH, RN; Constance M Johnson¹,⁴, PhD, RN, FAAN; Allison A Vorderstrasse⁵, DNSc, APRN, FAAN; Sahiti Myneni¹, PhD, MSE

¹School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
²Durham Veterans Affairs Medical Center, Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham, NC, United States
³Duke University School of Nursing, Durham, NC, United States
⁴Cizik School of Nursing, The University of Texas Health Science Center at Houston, Houston, TX, United States
⁵Rory Meyers School of Nursing, New York University, New York, NY, United States

Corresponding Author:
Carlos A Pérez-Aldana, MSc
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
Suite 600
7000 Fannin
Houston, TX, 77030
United States
Phone: 1 713 500 3900
Email: Carlos.A.PerezAldana@uth.tmc.edu

Abstract

Background: Diabetes remains a major health problem in the United States, affecting an estimated 10.5% of the population. Diabetes self-management interventions improve diabetes knowledge, self-management behaviors, and clinical outcomes. Widespread internet connectivity facilitates the use of eHealth interventions, which positively impacts knowledge, social support, and clinical and behavioral outcomes. In particular, diabetes interventions based on virtual environments have the potential to improve diabetes self-efficacy and support, while being highly feasible and usable. However, little is known about the patterns of social interactions and support taking place within type 2 diabetes-specific virtual communities.

Objective: The objective of this study was to examine social support exchanges from a type 2 diabetes self-management education and support intervention that was delivered via a virtual environment.

Methods: Data comprised virtual environment–mediated synchronous interactions among participants and between participants and providers from an intervention for type 2 diabetes self-management education and support. Network data derived from such social interactions were used to create networks to analyze patterns of social support exchange with the lens of social network analysis. Additionally, network correlations were used to explore associations between social support networks.

Results: The findings revealed structural differences between support networks, as well as key network characteristics of supportive interactions facilitated by the intervention. Emotional and appraisal support networks are the larger, most centralized, and most active networks, suggesting that virtual communities can be good sources for these types of support. In addition, appraisal and instrumental support networks are more connected, suggesting that members of virtual communities are more likely to engage in larger group interactions where these types of support can be exchanged. Lastly, network correlations suggest that participants who exchange emotional support are likely to exchange appraisal or instrumental support, and participants who exchange appraisal support are likely to exchange instrumental support.

Conclusions: Social interaction patterns from disease-specific virtual environments can be studied using a social network analysis approach to better understand the exchange of social support. Network data can provide valuable insights into the design of novel and effective eHealth interventions given the unique opportunity virtual environments have facilitating realistic environments that are effective and sustainable, where social interactions can be leveraged to achieve diverse health goals.

(JMIR Diabetes 2021;6(1):e21611) doi:10.2196/21611
KEYWORDS

type 2 diabetes; diabetes education; self-management; social support; virtual environments; social network analysis

Introduction

Overview

Diabetes remains a major health problem in the United States, affecting an estimated 34.2 million people of all ages (about 10.5% of the country’s population) [1]. Data show that type 2 diabetes (T2D) accounts for the most diabetes burden (between 90% and 95%), and its prevalence will continue to increase [1,2]. Diabetes is a challenging chronic illness because self-management is critical to reduce and delay the onset of complications and mortality [3-6]. Several evidence-based strategies, such as diabetes self-management education (DSME) and ongoing self-management support by peers and providers, have been shown to be effective in the management of T2D [7-9]. In particular, self-management is important in T2D given that patients manage 99% of their own care [10,11]. Moreover, diabetes self-management interventions improve diabetes knowledge and self-management behaviors, in addition to clinical outcomes [12]. Despite these benefits, less than 60% of people with diabetes attend DSME and only about 7% of newly diagnosed patients with diabetes attend DSME within 12 months following their diagnosis [13-16], indicating a pressing need for the delivery of accessible DSME and ongoing self-management support interventions.

Widespread internet connectivity provides new opportunities for wider web technology access and use by patients. Internet-based interventions, also known as eHealth, can connect patients to both peers and providers to facilitate support as well as access to evidence-based information [17]. Research suggests that T2D interventions incorporating interactive, individualized, and frequent interactions among patients, educators, and providers are among the most effective approaches [9]. eHealth interventions can provide such interactions in an effective and accessible way, which otherwise would be costly and unsustainable [12]. In addition, eHealth interventions have shown positive impacts on knowledge, social support, and clinical and behavioral outcomes [18]. Johnson et al have highlighted the benefits of eHealth interventions on T2D management, such as increased support, self-efficacy, and knowledge; improvements in glycemic levels and self-management behaviors; and efficient use of primary care services [12]. Furthermore, successful eHealth programs focused on DSME provided relevant content, engaging interactive elements, personalized learning experiences, and self-assessment tools for monitoring and feedback [17-20]. However, in spite of the potential benefits eHealth offers for DSME, eHealth interventions have been mostly based on traditional website formats. Such website formats generally lack realistic simulated environments where DSME actually takes place, such as patient community places (eg, grocery stores and restaurants) [7,21].

Virtual Environments and Diabetes Self-Management

Education and Support

Virtual environments offer an effective way to provide patients with realistic settings for the acquisition and application of knowledge in community settings where daily T2D self-management takes place, while addressing barriers such as transportation, cost, time, and scheduling issues [22]. In addition, virtual environments have started to show a potential to improve diabetes self-efficacy and social support, while being highly feasible and usable [12]. Second Life (Linden Lab), a highly popular virtual world, has been shown to be an effective tool that can lead to “significant learning gains” [23]. Second Life allows users to socialize and behave in a similar way as they would naturally do in normal settings through virtual human representations known as avatars [24]. Furthermore, virtual environments, such as Second Life, offer the potential for users to perform behaviors within realistic scenarios by providing them with presence, immersion, and social interaction, while facilitating communication between patients, educators, and providers [12,24]. While virtual environments have been used to deliver health information, education, social support, and social networking, most Second Life–based health sites to date have focused on disseminating information and offering support groups [24].

Self-management diabetes interventions based on virtual environments enable diabetes education, the development of new skills, and the exchange of peer support in synchronous and asynchronous ways [7]. The Second Life Impacts Diabetes Education & Self-Management (SLIDES) virtual community was among the first interventions aimed at providing DSME and support using Second Life [24]. The results of SLIDES showed improvements in diabetes self-efficacy, social support, and foot care, as well as trends toward improvements in diet, weight loss, and clinical outcomes, while being highly feasible and usable [12]. The development of the SLIDES platform, as well as its preliminary effects, is described elsewhere [12,24]. Virtual environments, such as SLIDES, are innovative ways to provide accessible DSME and ongoing self-management support. A key characteristic of these environments is the potential for participants to develop real-world skills via simulation and rehearsal within the virtual environment that can be transferable and thus affect behaviors in the real world [12].

Another significant characteristic of virtual environments is the facilitation of social support among participants [12,24]. Social support is generally described as “an exchange of resources between at least two persons aimed at increasing the wellbeing of the receiver” [25-27]. Social support is recognized as a key component of diabetes self-management, in addition to adequate skills and behavioral development [22,28,29]. Studies have shown that social support is commonly provided through social interactions to achieve health outcomes [30,31]. Moreover, research suggests that people with T2D can benefit from frequent and sustained social interactions among peers and providers by obtaining education and support [28,32-34]. In addition, T2D interventions that are based on virtual environments can provide realistic, personalized, and ongoing interaction and support that assist participants in health care decision making [7,12,34-36]. SLIDES showed that virtual
environment-mediated interactions resemble physical ones; therefore, patients with T2D are presented with the possibility of greatly improving their access to social support [12,34]. However, the social networks highlighting the patterns of interactions within T2D-specific virtual communities, such as SLIDES, have not been studied. While the prominent effects of social relationships on health decisions and related behavior changes have been established [37,38], little is known about social interactions and the exchange of support in disease-specific virtual environments.

Social Network Analysis and Online Health Communities

The study of social networks provides researchers with a unique opportunity to get an in-depth view and a better understanding of the structure of online communities [38,39]. Social network research has shown that social connections (i.e., peers, family members, etc.) disseminate health information, provide social support, and influence health behaviors [38,39]. Social network analysis (SNA) has been used to study the ways in which social connections can influence individuals’ attitudes, beliefs, and behaviors. Such network influences can be caused by the network environment, the position an individual occupies in the network, or structural or network-level properties [38,39]. For example, being central in a social network determines a high importance for information dissemination. Similarly, individuals located on a network’s periphery, known as peripheral individuals, can act as bridges connecting otherwise disconnected groups, thus enabling collective actions. Peripheral individuals are characterized by having one or few connections on the outside of a network and thus participating infrequently. Moreover, peripheral individuals are usually free from social norms and constraints, and thus, innovation can occur [38,39]. Furthermore, network structural properties, such as clustering, can help to identify highly connected groups of individuals, where behavior change can be accelerated. Lastly, densely connected networks have been shown to generate faster diffusion and increased coordinated action [38,39].

SNA is increasingly becoming useful to the study of online health communities owing to the exponential growth in the use of electronic communications [40]. The massive amounts of social interactions taking place within online communities today are providing researchers with valuable network data. Research has focused on the analysis of online social interactions from both general purpose social media platforms (e.g., Twitter and YouTube) and health care-specific platforms (e.g., American Diabetes Association online community) [41-44]. Often, qualitative analysis and computational text analysis are used to analyze social media interactions [41-43]. Studies have shown that SNA provides insights into social influence, information dissemination, and behavioral diffusion [39,40,45,46]. On one hand, communication structure (who communicates with whom) is key for the study of peer influence on health behaviors [40]. On the other hand, analyses of the structures of online peer-to-peer communications provide valuable insights into opinion leaders [40,45,47]. Both approaches have the potential to help researchers model effective network data–based interventions [40]. Similarly, social support exchange patterns within disease-specific virtual communities, such as SLIDES, can be studied using a SNA approach, which would allow the visualization and description of communication structures, peer influences, and behavioral diffusion, as well as the impact on health outcomes, such as blood glucose levels, for patients with diabetes [45-50]. However, despite the benefits SNA offers, to our knowledge, social interactions occurring within virtual environments have not been studied using this approach. In this study, a secondary data analysis of SLIDES social interactions through the SNA lens was carried out to examine social support exchange patterns between participants and providers [12,24,34].

Research Aims

The overall goal of our study was to examine social support exchanges from a T2D self-management education and support intervention (SLIDES) that was delivered via a virtual environment. The specific aims of our study were as follows: (1) to examine patterns of social interaction and support of the SLIDES intervention by creating network structures for different types of social supports and assessing these support networks using quantitative network measures; (2) to explore the associations between social support network structures by correlating them with each other using the quadratic assignment procedure (QAP); and (3) to provide insights into the exchange of social support within a disease-specific virtual environment.

Methods

SNA Methodology

Social Network Data

SLIDES social interaction data were used for our study [34]. SLIDES included a total sample of 24 individuals, with 20 participants and 4 providers (including diabetes educators and moderators). Detailed participant demographics are described elsewhere [12]. SLIDES facilitated virtual interactions among participants with T2D and providers in the following two types of sessions: education and support. Education sessions were held twice a week, and support sessions were held weekly. SLIDES social interactions consisted mostly of synchronous naturalistic conversations that took place throughout different locations within the virtual environment (e.g., bookstore, restaurant, and classroom) [12,24]. These conversations enabled the exchange of social support among participants and between participants and providers, and were continuously recorded and transcribed [12,24]. These transcriptions provided the data set from which network data were derived for our analysis. Detailed information on the SLIDES study site, theoretical framework, sample, measures, and outcomes have been published elsewhere [12,24]. Our analysis focused on interactions where social support was exchanged among participants and between participants and providers during a 6-month study enrollment period [34]. Study participants could log into SLIDES and participate as much or as little as they wanted and engage in synchronous conversations. Social support was defined as “personal informal advice and knowledge that help individuals initiate and sustain T2D self-management behaviors, thus increasing adherence” [22,25,27,30,34]. Social support types included emotional, instrumental, informational, and appraisal [22,25-27,29,34]. SLIDES social interactions, which were
previously characterized by the aforementioned types of social support [34,51], were used to create network structures in order to analyze social support exchange patterns at the group level (ie, participants/providers who interacted in a conversation by either listening or engaging directly, where a certain type of support was exchanged, were all linked together for that particular conversation). Thus, the unit of analysis included the tie among participants and between participants and providers who interacted via synchronous conversations, as well as the types of social support exchanged in each transcribed conversation as previously characterized [34,51].

### Network Structures and Measures

Network structures were created for each type of social support by representing participants and providers as nodes and representing interactions where social support was exchanged as edges (interconnections between nodes). For each type of social support network, all edges indicating who participated in a conversation were included (ie, who interacted with whom during a virtual conversation in which social support was exchanged). Quantitative network measures were used to assess network structures across all types of social support. Network measures explain structural differences (eg, density and cohesion), as well as node importance within a network (eg, centrality) [38,39]. The following network measures were used: average degree (average number of connections of all nodes; a higher average degree number means that members of a network interacted with a higher number of members via synchronous conversations, either on a one-to-one basis or at a group level); graph density (proportion of connections relative to the total number of possible connections; ranging from 0 to 1; a higher graph density means that members of a network most likely engaged in conversations involving a higher number of members, ie, larger groups); average path length (average distance between all node dyads; the distance of a dyad is 1, which means a direct interaction between two members of the network; a higher average path length is associated with a higher distance or number of steps required for two network members to interact with each other, resulting in a less efficient network); average clustering coefficient (average measure of the interconnectivity of the node neighborhood; ranging from 0 to 1; a higher average clustering coefficient means that node neighborhoods are more interconnected, indicating conversations among a larger number of members for larger node neighborhoods); and modularity (the level of development of subcommunities within a network; ranging from −1 to 1; higher modularity values indicate higher levels of subcommunity development within a network) [38,39].

### Network Statistical Analysis

Once network structures were created, we correlated them with each other to explore associations between social support network structures. The QAP was used to test network correlations. QAP is a nonparametric method based on permutations that allows testing structural similarities (correlations) between social network structures [52]. We used Gephi version 0.9.2 and Ucinet version 6.685 (Analytic Technologies) to create network structures and to calculate network measures, as well as to perform correlation analysis [53,54].

**Results**

**Network Structures**

Figure 1 shows a network structure depicting all SLIDES social interactions where all types of social support were exchanged among participants and between participants and providers. Network structures for each type of social support exchanged by SLIDES participants are shown in Figure 2.

**Figure 1.** Network structure of social interactions where all types of social supports were exchanged. Node size indicates degree and node color indicates the existence of three subcommunities or groups, with one larger subcommunity shown in orange and two smaller subcommunities shown in purple and grey. Further, edge thickness represents the frequency of interactions when members communicated more often.
Figure 2. Network structures of Second Life Impacts Diabetes Education & Self-Management (SLIDES) social support interactions by the type of support. Node size indicates degree and node color indicates the existence of subcommunities, where larger subcommunities are shown in orange and smaller subcommunities are shown in purple and grey.

In addition, Table 1 summarizes the network measures for each social support network. As seen in Figure 2, the emotional and appraisal support networks were the most populous, with the former comprising 24 nodes and 1219 edges and the latter comprising 20 nodes and 737 edges. Moreover, the emotional and appraisal support networks had the highest average degrees (9.08 and 9.5, respectively) compared with the instrumental and informational support networks (6.0 and 3.2, respectively). This indicates that each member of these support networks interacted on average with nine other members via synchronous conversations, either on a one-to-one basis or at a group level, thus making them the most active networks. Additionally, assessment of degree at a node level showed that all support networks were somewhat centralized around a few nodes, suggesting that some members were more popular. Furthermore, the appraisal (0.5) and instrumental (0.43) support networks were the densest, suggesting that members of these networks most likely engaged in conversations involving a higher number of members (ie, larger groups), where some participants directly exchanged appraisal and/or instrumental support, while other members of the group had a latent exposure to this support.
Additionally, no substantial differences were observed between all average path length values. However, the appraisal (1.52) and instrumental (1.62) support networks had a slightly lower average path length compared with the emotional (1.74) and informational (1.98) support networks. This indicates that the distance or number of steps needed for members of these networks to interact with each other required on average fewer steps to exchange the supports, thus making these networks more efficient. In terms of network structure and community development, on one hand, the instrumental, emotional, and appraisal support networks had higher average clustering coefficients (76%, 73%, and 72%, respectively) compared with the informational support network (57%). These results indicate high levels of interconnectivity within these support networks. On the other hand, the modularity values of the emotional (0.11), appraisal (0.12), and instrumental (0.12) support networks were lower compared with that of the informational (0.46) support network. This indicates that subcommunities of network members exchanging informational support reached higher levels of development in comparison with subcommunities from all other support networks.

Lastly, Figure 3 illustrates a two-mode network representing the affiliation between participants and providers, and the types of social support exchanged via social interactions. As seen in Figure 3, according to degree, the two-mode network is centralized around emotional and appraisal support, indicating that a higher number of participants and providers participated in interactions where these types of support were exchanged (either directly or indirectly having a latent exposure as previously discussed). Moreover, a subgroup of participants and providers engaged more frequently in interactions where emotional support and appraisal support were exchanged, which are represented by thicker edges.

Figure 3. Two-mode network structure of social interactions for all types of support. The shape of the nodes distinguishes two sets of nodes as follows: squares represent participants and providers, and circles represent types of social support. In addition, the color of the circles represents each type of social support (orange, purple, yellow, and blue representing emotional, appraisal, informational, and instrumental support, respectively). Finally, the size of the circles indicates degree, and edge thickness represents the frequency of participants’ interactions within each type of support.

Network Statistical Analysis

Table 2 shows network correlation scores obtained by QAP analysis. All social support networks were correlated with one another. QAP correlation scores between the emotional and appraisal, informational and appraisal, and instrumental and informational support networks were much stronger when compared with the correlations between the informational and appraisal, informational and emotional, and instrumental and informational support networks. The stronger correlation scores suggest that considerable similarities exist between the aforementioned social support networks.

Table 1. Summary of social network metrics for Second Life Impacts Diabetes Education & Self-Management (SLIDES) social support networks.

<table>
<thead>
<tr>
<th>Social support network</th>
<th>Average degree</th>
<th>Graph density</th>
<th>Average path length</th>
<th>Clustering coefficient</th>
<th>Modularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional</td>
<td>9.08</td>
<td>0.39</td>
<td>1.74</td>
<td>0.73</td>
<td>0.11</td>
</tr>
<tr>
<td>Instrumental</td>
<td>6.0</td>
<td>0.43</td>
<td>1.62</td>
<td>0.76</td>
<td>0.12</td>
</tr>
<tr>
<td>Informational</td>
<td>3.2</td>
<td>0.35</td>
<td>1.98</td>
<td>0.57</td>
<td>0.46</td>
</tr>
<tr>
<td>Appraisal</td>
<td>9.5</td>
<td>0.5</td>
<td>1.52</td>
<td>0.72</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Table 2. Network correlation test results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Appraisal</th>
<th>Emotional</th>
<th>Informational</th>
<th>Instrumental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>1</td>
<td>0.974</td>
<td>0.344</td>
<td>0.833</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.004</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Emotional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>0.974</td>
<td>1</td>
<td>0.318</td>
<td>0.818</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>—</td>
<td>.003</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Informational</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>0.344</td>
<td>0.318</td>
<td>1</td>
<td>0.204</td>
</tr>
<tr>
<td>P value</td>
<td>.004</td>
<td>.003</td>
<td>—</td>
<td>.02</td>
</tr>
<tr>
<td>Instrumental</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>0.833</td>
<td>0.818</td>
<td>0.204</td>
<td>1</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.02</td>
<td>—</td>
</tr>
</tbody>
</table>

aNot applicable.

Discussion

Principal Findings

In this study, we used SNA to examine patterns of social interactions and support of SLIDES, an intervention for T2D self-management education and support that was delivered via a virtual environment [12,24]. To the best of our knowledge, this study is among the first to explore the patterns of social interactions of a disease-specific virtual environment. This novel approach provided insights into the exchange of social support within the SLIDES virtual community. Our findings indicate that emotional and appraisal support networks were the largest, most centralized, and most active, indicating that a virtual community with a larger number of members can be more supportive. Moreover, a higher centralization indicated that some network members were more active, which suggests that a virtual community benefits from having active members, such as educators and moderators, because they can help engage the community. This is important for the design of interventions based on virtual environments. For example, interventions could recruit diabetes moderators or leaders to act as peer influencers or change agents. Moreover, appraisal and instrumental support networks are more connected than emotional and informational support networks. This suggests that more members are likely to engage in larger group synchronous conversations, thus indicating that well-connected networks can facilitate the exchange of appraisal and instrumental support within virtual communities. This finding could be leveraged when designing interventions that facilitate the exchange of appraisal and/or instrumental support.

An analysis of the structures of the support networks revealed higher levels of interconnectivity within the instrumental, emotional, and appraisal support networks, as indicated by their higher average clustering coefficients. Clustering can accelerate information and behavior spread [38,39], thus suggesting that interventions based on virtual environments can leverage this characteristic to accelerate the exchange of social support. Despite high degrees of clustering, instrumental, emotional, and appraisal support networks had low modularity values, indicating low levels of subcommunity development. In contrast, the informational support network showed a higher level of subcommunity development. From an intervention’s perspective, subcommunities or groups within informational support networks can be leveraged to spread resources and behaviors, in addition to providing informational support. Studies have shown that groups have norms and exert social pressure, enabling behavior change, as well as more opportunities to access information, resources, and support [39].

Our findings also show that a higher number of participants and providers participated in interactions where emotional support and appraisal support were exchanged, and they did so more frequently. These findings diverge from a previous analysis by Lewinski et al, where informational support and emotional support were the most commonly exchanged types of support among participants and between participants and providers, and appraisal support exchange was lower [34]. Their analysis focused on support exchanges at a dyadic level in order to characterize interactions. In contrast, our analysis focused on support exchanges at a group level, as previously indicated. In other words, a dyadic analysis for two participants who interact in a group conversation would identify the frequency of support exchanged between those two participants. On the other hand, our network approach to this same scenario would take into account the connections between all participants who engaged in the conversation, including those who actively engaged one another to exchange support, as well as the other participants who engaged passively and had a latent exposure. Taking this into account, we hypothesize that a higher and more frequent engagement in interactions where emotional and appraisal support were exchanged was caused by the role providers, specifically diabetes educators, played assisting in the self-management of diabetes.

Lastly, network correlations showed that all social support networks were correlated with one another. Specifically, stronger
correlation scores for emotional and appraisal, instrumental and appraisal, and instrumental and emotional support networks indicate that considerable similarities exist between these networks. These results suggest that SLIDES participants who exchanged emotional support were likely to exchange appraisal or instrumental support. Likewise, participants who exchanged appraisal support were likely to exchange instrumental support. From an intervention’s perspective, educators and moderators from virtual communities can leverage interactions where a certain type of support is exchanged in order to maximize the provision of advice and support among members of such communities. For example, by promoting interactions between members where emotional support is exchanged, further discussion and opportunities could be created that would most likely prompt exchange of appraisal or instrumental support [34,55,56]. As a result, a higher number of supportive relationships would be fostered among participants and providers, increasing the effectiveness of support networks and thus substantiating the value of virtual communities for diabetes self-management and other health goals.

Limitations
There are several limitations in this study. The small sample size of the SLIDES study (N=24) created a small virtual community, which consequently resulted in a small community. The social dynamics resulting from a small community might differ from larger ones, which suggests that our findings should be interpreted with caution. The creation of social networks from interactions, where some type of social support was exchanged, was considered at a group conversational level and not at a dyadic level. This resulted in group identification of social support interactions, meaning that a type of social support was assigned to all group participants interacting in a conversation where social support occurred during a particular conversation. Future studies could improve network creation by analyzing participants’ interactions at a dyadic level so that social support exchanges describe social ties at a dyadic level, thus providing more accurate social support dynamics. Despite these limitations, we consider these findings valuable because of the insights provided into social support exchanges within disease-specific virtual environments.

Conclusions
This study described the utility of SNA to examine social support in a DSME virtual environment. Our findings have revealed structural differences between support networks, as well as key network characteristics of supportive interactions facilitated by the virtual community, with emotional and appraisal networks being large, centralized, and most active, thus emphasizing the value of virtual environments as sources of these two support types for T2D patients. In addition, support networks have highlighted the benefits of central members, such as educators and moderators, who can contribute by facilitating community engagement. Specifically, educators and moderators from the SLIDES intervention have facilitated community engagement by leading weekly synchronous group meetings that include educational sessions, focusing on core American Diabetes Association/American Association of Diabetics Education self-management curriculum, as well as support sessions [12].

Furthermore, our appraisal and instrumental support networks suggest that members of virtual communities are more likely to engage in larger group interactions where these types of support can be exchanged, with the caveat that some members can engage one another to actively exchange support, while the other members engage passively and have a latent exposure to support exchange. Lastly, our network correlation analysis has shown that participants who exchange emotional support are likely to exchange appraisal or instrumental support, and participants who exchange appraisal support are likely to exchange instrumental support. These associations suggest that interactions, where a certain type of support is exchanged, could be leveraged to maximize the provision of advice and support among network members, thus increasing the effectiveness of support networks enabled by virtual communities.

Network data can provide valuable insights into the design of novel and effective digital health interventions given the unique opportunity disease-specific virtual environments have for facilitating realistic environments that are effective and sustainable, where social interactions can be leveraged to achieve diverse health goals.

Acknowledgments
Data in this study were obtained in the following grants: F31-NR016622-01 (principal investigator [PI]: Lewinski) funded by the National Institutes of Health, National Institute for Nursing Research and 1R21LM010727-01 (PI: Johnson) funded by the National Library of Medicine. Support for Dr Lewinski was provided by the VA Office of Academic Affiliations (TPH 21-000), and publication support was provided by Durham VA Health Services Research Center of Innovation funding (CIN 13-410). Part of the research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under Award Number 1R01LM012974-01A1. The findings and conclusions in this document are those of the authors who are responsible for its contents and do not represent the views of the Department of Veterans Affairs or the National Institutes of Health; therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs.

Conflicts of Interest
AAL reports receiving funds from PhRMA Foundation and Otsuka. Other authors have no conflicts to declare.

References
http://diabetes.jmir.org/2021/1/e21611/

JMIR Diabetes 2021 | vol. 6 | iss. 1 | e21611 | p.83
(page number not for citation purposes)
Abbreviations

- **DSME**: diabetes self-management education
- **QAP**: quadratic assignment procedure
- **SLIDES**: Second Life Impacts Diabetes Education & Self-Management
- **SNA**: social network analysis
- **T2D**: type 2 diabetes

©Carlos A Pérez-Aldana, Allison A Lewinski, Constance M Johnson, Allison A Vorderstrasse, Sahiti Myneni. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 25.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Early Insights From a Digitally Enhanced Diabetes Self-Management Education and Support Program: Single-Arm Nonrandomized Trial

Folasade Wilson-Anumudu¹, MPH; Ryan Quan¹, MPH; Cynthia Castro Sweet¹, PhD; Christian Cerrada², PhD; Jessie Juusola², PhD; Michael Turken¹, MD, MPH; Carolyn Bradner Jasik¹, MD

¹Omada Health, Inc, San Francisco, CA, United States
²Evidation Health, Inc, San Mateo, CA, United States

Corresponding Author:
Folasade Wilson-Anumudu, MPH
Omada Health, Inc
500 Sansome Street, Suite 200
San Francisco, CA, 94111
United States
Phone: 1 6502696532
Email: folasade.anumudu@omadahealth.com

Abstract

Background: Translation of diabetes self-management education and support (DSMES) into a digital format can improve access, but few digital programs have demonstrated outcomes using rigorous evaluation metrics.

Objective: The aim of this study was to evaluate the impact of a digital DSMES program on hemoglobin A₁c (HbA₁c) for people with type 2 diabetes.

Methods: A single-arm, nonrandomized trial was performed to evaluate a digital DSMES program that includes remote monitoring and lifestyle change, in addition to comprehensive diabetes education staffed by a diabetes specialist. A sample of 195 participants were recruited using an online research platform (Achievement Studies, Evidation Health Inc). The primary outcome was change in laboratory-tested HbA₁c from baseline to 4 months, and secondary outcomes included change in lipids, diabetes distress, and medication adherence.

Results: At baseline, participants had a mean HbA₁c of 8.9% (SD 1.9) and mean BMI of 37.5 kg/m² (SD 8.3). The average age was 45.1 years (SD 8.9), 70% were women, and 67% were White. At 4-month follow up, the HbA₁c decreased by 0.8% (P<.001, 95% CI –1.1 to –0.5) for the total population and decreased by 1.4% (P<.001, 95% CI –1.8 to –0.9) for those with an HbA₁c of >9.0% at baseline. Diabetes distress and medication adherence were also significantly improved between baseline and follow up.

Conclusions: This study provides early evidence that a digitally enhanced DSMES program improves HbA₁c and disease self-management outcomes.

(JMIR Diabetes 2021;6(1):e25295) doi:10.2196/25295

KEYWORDS
diabetes education; digital health; remote monitoring; type 2 diabetes

Introduction

Background

Over 34 million people in the United States have diabetes (9% of the adult population), and 1 in 4 health care dollars spent in the United States is for diabetes care [1]. Among all diabetes cases, 90%-95% are type 2 diabetes mellitus (T2DM) [2]. A core component of diabetes management is comprehensive diabetes self-management education and support (DSMES), which is associated with improved outcomes and lower costs [3-5]. DSMES is traditionally delivered in person, either one on one or in a group setting with a certified diabetes care and education specialist (CDCES).
DSMES is widely covered by private and public insurance, including Medicare, and is typically prescribed by a physician at diagnosis, when education gaps exist, or when the treatment plan is changed. The primary goal of DSMES is to help patients acquire the knowledge, skills, and abilities for diabetes self-care [6]. Core educational topics include disease awareness, glucose monitoring, medication adherence, nutrition support, delay of complications, and problem-solving [7].

Despite the widely accepted benefits of DSMES, access remains a challenge. Only 43 states and 57% of counties in those states have accredited DSMES programs in the United States [8]. As of 2017, only 52% of people diagnosed with diabetes in the United States have accessed self-management support services, with rates decreasing in recent years [9]. To address the unmet need, technology-enabled platforms have emerged as a more accessible venue for DSMES delivery. There are numerous commercial products available that allow people to access DSMES programs through personal mobile devices (eg, smartphones, tablets, laptops) with a wide range of approaches [10,11]. Staffing varies widely from none (100% patient-driven) to uncredentialed coaches to CDCES.

Technology-based DSMES programs have demonstrated a positive impact on hemoglobin A1c (HbA1c) in academic settings with noncommercially available programs [12]. These interventions typically adhere to DSMES guidelines and include credentialed staff for program delivery. Commercially available technology-based DSMES solutions in the market are often limited by lack of accreditation, uncredentialed staff, and research results produced from less rigorous methods [13]. Although some studies have demonstrated that commercially available DSMES programs improve diabetes-related outcomes for users, the staffing, number of touchpoints, manner of delivery (asynchronous vs synchronous), and inclusion of connected devices, among other factors, vary widely among programs [14-16]. As such, more research is needed to understand best practices for digital DSMES delivery. Furthermore, methodologically rigorous research is also needed to demonstrate the parity of outcomes to in-person care [12].

Objective
The goal of this pilot study was to evaluate the impact of a digital DSMES program enhanced with deep lifestyle and behavior change support on HbA1c for people with T2DM and elevated HbA1c. We hypothesized that the digital DSMES program would be associated with greater improvements in HbA1c for people who were furthest away from their HbA1c goal (baseline HbA1c ≥ 9.0%) at the start of the program. We further evaluated the impact of the digital DSMES program on cardiovascular and patient-reported outcomes, as cardiovascular risk factors are a frequent comorbidity of diabetes.

Methods
Participants
We invited members of an online health community to participate in this study (Achievement, Evidation Health Inc). Achievement is a web- and mobile-based community in the United States where members can connect their activity trackers, and fitness and health apps to the platform and, by logging activities, accumulate points that are redeemable for monetary rewards. Additionally, members self-report on various health conditions and are invited to participate in remote research opportunities as relevant studies become available. In this study, recruitment was targeted to members who had self-reported a diagnosis of T2DM. Invited members were linked to an online research study platform (Achievement Studies, Evidation Health Inc) where study eligibility was assessed using automated screener questions. Individuals who lived in the United States, were at least 18 years of age, self-reported a T2DM diagnosis, self-reported HbA1c of 7.5% or greater, had a BMI ≥ 25 kg/m² (≥23 kg/m² if they self-identified as Asian), and had access to a computer or smartphone to participate in the digital DSMES program were eligible for the study.

Procedures
If deemed eligible after completing the screener, potential participants continued in the online study platform to sign an electronic informed consent form and completed an online baseline survey, which consisted of questions about their demographics, health and diabetes history, and patient-reported outcomes. They then completed a baseline visit at a Quest Diagnostics Patient Service Center (PSC) of their choosing. The baseline visit consisted of a venous whole blood draw, physical measurements (height, weight, waist circumference), resting blood pressure, and resting heart rate. After completing the PSC visit, potential participants were instructed to set up their account on the digital DSMES program. After completion of signed electronic informed consent form, and the PSC visit and program account setup, individuals were considered enrolled in the study. Participants were able to reach out to research staff with questions via email or phone through the online study platform before and during the enrollment process, and could continue to reach out throughout the study.

During the study period, participants were encouraged to engage with the DSMES program. All participants were provided a cellularly connected weight scale that was linked to their program account. Participants who were advised to use monitoring devices in their diabetes self-care were provided cellularly connected blood pressure monitors and glucose meters. Participants were also able to access their personal online study platform dashboard to complete study procedures and keep track of their progress throughout the study through the use of any web-enabled device. Approximately 4 months after enrollment, participants repeated the online survey and clinical outcome measures (HbA1c, blood pressure). Participants received compensation for completing each study-related task such as surveys and lab visits. This study was approved by the Western Institutional Review Board (Puyallup, WA).

Study Outcomes
The primary outcome of this study was change in HbA1c from baseline to 4 months, as well as changes in HbA1c based on starting HbA1c values. Secondary outcomes included changes in cardiovascular risk factors (blood pressure, total cholesterol [TC]) among those who started the study with elevated risk values.
Factors, in addition to changes in diabetes distress and medication adherence from baseline to 4 months.

**Measurements**

At baseline, participants completed an assessment at the PSC that included 13 mL venous whole blood specimen collection under sterile conditions by a trained phlebotomist. The nonfasting blood specimens were processed for HbA1c and a lipids panel (TC, high- and low-density lipoprotein [HDL, LDL], and TC/HDL ratio). A trained technician collected blood pressure after a 5-minute quiet resting period with legs uncrossed using an automatic blood pressure monitor and size-adjustable cuff. Height was measured to the nearest centimeter using a calibrated stadiometer with the participant in stocking feet. Weight was measured using a calibrated scale with the participant in light clothing and no shoes. Waist circumference was measured in whole units (inches) using a nonstretachable measuring tape above the first layer of clothing. BMI was calculated from weight in kilograms divided by height in meters squared. Results were sent by Quest Diagnostics and accessed by the research team via secure file transfer. Participants received copies of their results both via secure email and mail.

Participants completed an online survey of patient-reported outcomes including the Diabets Distress Scale (DDS), a 17-item scale of different dimensions of distress and burden related to diabetes, which has been shown to have reliability and validity [17], and the Simplified Medication Adherence Questionnaire (SMAQ), a 6-item measure that categorizes respondents as adherent or nonadherent based on recent patterns of medication-taking behaviors [18].

The original protocol planned for a repeat assessment using identical methods 4 months after enrollment. However, the 4-month assessments were scheduled to begin in April of 2020, during the height of the COVID-19 pandemic [19]. People with diabetes are at high risk for severe illness from COVID-19 [20]; therefore, the study protocol was changed to eliminate the in-person visit to support participants to shelter in place. In replacement of the venipuncture blood draw, a Quest Diagnostics Qcard self-collection card was sent to each participant for collection of HbA1c and blood lipids data. The Qcard is a self-collection card that uses the dried blood spot method, with a correlation to venipuncture HbA1c in the range of 0.95 to 1.0 [21]. Triglycerides and LDL were not available through the Qcard and as such were removed as study outcomes. Weight at the 4-month time point was collected using a calibrated stadiometer with the participant in stocking feet, BMI was calculated from weight in kilograms divided by height in meters squared. Results were sent by Quest Diagnostics and accessed by the research team via secure file transfer. Participants received copies of their results both via secure email and mail.

Participants were assigned to a CDCES who provided individualized coaching around the American Association of Diabetes Educators 7 self-care behaviors [22]. They were also placed in a virtual peer group including other program participants with T2DM, and could communicate with peers through a secure discussion board. As needed, the CDCES referred participants back to their primary care team for medication reviews or adjustments as their health targets and self-care goals were achieved. The program is accredited by the American Association of Diabetes Care and Education Specialists [23]. The program takes a user-centered approach that encourages participants to engage at a time and frequency they choose, and with the tools and resources they find most useful, and does not have any predetermined volume or pattern that participants are expected to engage in program features.

**Intervention**

Omada for Diabetes is a digitally enhanced DSMES program designed to build self-management skills and support diabetes management between outpatient visits with primary care providers and specialists to ensure that users achieve their health targets (eg, HbA1c, blood pressure, cholesterol) and obtain health maintenance services (eg, screening for neuropathy and retinopathy). The program offers disease education, comprehensive lifestyle self-management support (ie, support for weight loss, dietary changes, physical activity increases), support for involvement in members’ current medication regimen, and support for use of monitors or trackers for their blood sugar and blood pressure, which are often used to inform small modifications in food intake, physical activity, medication, or communication with health care providers. Participants used a technology-enabled platform with a portable interface to a variety of personal mobile devices. All participants received a cellurally connected BodyTrace weight scale, and if needed, a blood glucose monitor (3G BioTel Care, Telcare LLC, Concord, MA) was also provided. Participants were assigned to a CDCES who provided individualized coaching around the American Association of Diabetes Educators 7 self-care behaviors [22].

The study was powered to detect a clinically meaningful 0.5% reduction in the primary outcome of HbA1c. With an estimated standard deviation of 1.8 and power set to 90%, the minimal sample size needed was 162. To allow for potential 20% loss to follow up and 10% of lab HbA1c values being below 7.5% at baseline, a total of 186 participants were planned for enrollment.

**Statistical Analysis**

The study was powered to detect a clinically meaningful 0.5% reduction in the primary outcome of HbA1c. With an estimated standard deviation of 1.8 and power set to 90%, the minimal sample size needed was 162. To allow for potential 20% loss to follow up and 10% of lab HbA1c values being below 7.5% at baseline, a total of 186 participants were planned for enrollment.

Descriptive statistics are presented to describe the demographics and baseline health status of participants. Baseline correlations using Pearson and Spearman correlation coefficients were examined to determine variables (age, gender, BMI) that could potentially confound HbA1c outcomes. No significant correlations were detected; therefore, paired t tests were used to examine baseline to post-test differences in study outcomes. Post hoc analyses were performed to examine the change in HbA1c based on the starting HbA1c range, with the hypothesis that those with higher blood glucose levels may receive greater benefit. Elevated blood pressure and blood lipids were not among the criteria for study inclusion and were therefore assessed as secondary outcomes of interest; we examined changes specifically among those who began the study with elevated cardiovascular risk factors. The McNemar test was performed to examine the change in the proportion of the population that was adherent to medications from baseline to post-test. Program engagement is summarized using averages...
across several metrics to reflect how participants engaged with the program over the course of the 4-month study.

We analyzed outcomes using complete case analysis for those who returned 4-month clinical and patient-reported survey data. Using multiple imputation, with an imputation of baseline values for primary and secondary outcomes for those with missing data at 4 months, we found that outcomes were similar in magnitude and statistical significance using both analytic methods. Therefore, we present our findings on the sample using results from the complete case analysis.

Results

Study Recruitment

Although the recruitment goal was 162 participants with starting HbA$_{1c}$ above 7.5%, 32 of the first 100 participants’ laboratory HbA$_{1c}$ result was below the 7.5% threshold. Therefore, we changed the protocol to use the baseline HbA$_{1c}$ as a clinical criterion for the study and only accepted those with a lab HbA$_{1c}$ value of 7.5% or greater. We continued enrollment until we reached at least 162 participants with a baseline HbA$_{1c}$ of 7.5% or greater and allowed the 32 participants with a baseline HbA$_{1c}$ below 7.5% to remain in the study. The final enrolled sample was 195, including 163 with a baseline HbA$_{1c}$ of 7.5% or greater and 32 with a baseline HbA$_{1c}$ of less than 7.5%. Six participants were withdrawn from the study: 4 developed a medical condition that precluded participation and 2 requested to voluntarily withdraw. At post-test, 78.8% (n=149) of the remaining 189 participants completed the home test kit; 8 were not sent kits as they resided in states where the home test is not authorized for distribution, and 88.4% (n=167) completed the online questionnaire. Study completion was defined as a final HbA$_{1c}$ value or completion of the final online questionnaire. We compared baseline demographic and clinical values for participants who completed the 4-month data collection and those who were lost to follow up, and found no significant differences across any baseline characteristics. We define loss to follow up as incompletion of the primary outcome of HbA$_{1c}$.

See Figure 1 for the flow of participants through each stage of the study.

Figure 1. Study participant flowchart. HbA$_{1c}$: hemoglobin A$_{1c}$.
Participant Characteristics at Baseline

Baseline characteristics of participants are shown in Table 1. The average starting HbA$_{1c}$ was 8.9%; 50% began the study with an HbA$_{1c}$ of 9.0% or higher. The mean age was 45.1 years, and the majority of participants were female and White. On average, total cholesterol was in the normal range, and blood pressure was close to the nationally recommended goal for those with diabetes. As measured by the SMAQ, 19% of participants were adherent to their current medication regimen. The mean DDS score at baseline was 2.7. A total or subscale score $>2.0$ (moderate distress) is considered clinically meaningful; average scores $<2.0$ reflect little or no distress, between 2.0 and 2.9 reflect moderate distress, and $\geq 3.0$ reflect high distress [24].

Table 1. Baseline participant characteristics (N=195).

<table>
<thead>
<tr>
<th>Baseline characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>45.1 (8.9)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>136 (69.7)</td>
</tr>
<tr>
<td>Race/ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>131 (67.2)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>32 (16.4)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>17 (8.7)</td>
</tr>
<tr>
<td>Asian</td>
<td>6 (3.1)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>2 (1.0)</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.1)</td>
</tr>
<tr>
<td>BMI, mean (SD)</td>
<td>37.5 (8.3)</td>
</tr>
<tr>
<td>Weight (pounds), mean (SD)</td>
<td>235.6 (57.3)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>106.9 (26.0)</td>
</tr>
<tr>
<td>Hemoglobin A$_{1c}$, mean (SD)</td>
<td>8.9 (1.9)</td>
</tr>
<tr>
<td>Total cholesterol (mg/dL), mean (SD)</td>
<td>178.9 (43.3)</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg), mean (SD)</td>
<td>127.0 (16.1)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg), mean (SD)</td>
<td>82.0 (10.4)</td>
</tr>
<tr>
<td>Diabetes Distress Score, mean (SD)</td>
<td>2.7 (1.0)</td>
</tr>
<tr>
<td>Adherent to current medications, n (%)</td>
<td>36 (18.5)</td>
</tr>
</tbody>
</table>

$^a$There were no statistically significant differences across baseline characteristics among those with and without follow-up data.

Program Engagement

Averaged across the 16 program weeks, participants used their blood glucose meter an average of 7.4 times per week. Participants weighed in an average of 4.9 times per week, interacted with their CDCES an average of 1.6 times per week, completed an average of 0.8 lessons per week, interacted with their peer groups an average of 0.9 times per week, tracked their physical activity 5.3 times per week, and tracked meals an average of 10.2 times per week.

Diabetes Outcomes

Baseline to post-test changes in all study outcomes are shown in Table 2. Among all participants who completed both a baseline and 4-month HbA$_{1c}$ test (n=149), participants achieved a statistically significant decrease in HbA$_{1c}$ of 0.8% ($t_{148}=-6.2, P<.001$). Table 3 shows changes based on starting HbA$_{1c}$ values. Those who started the study with an HbA$_{1c}$ of 9.0% or higher saw the greatest magnitude of change, with an average decrease of 1.4% ($t_{72}=-6.1, P<.001$). Across the total sample, weight significantly decreased an average of 3.0 pounds over 4 months ($t_{146}=-2.2, P=.03$), and 18.4% of the sample achieved significant weight loss (>5% body weight) (Table 2).
Table 2. Baseline to post-test changes in clinical outcomes (N=167).

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>n</th>
<th>Baseline</th>
<th>Post-test</th>
<th>Difference</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HbA\textsubscript{1c} (%)</td>
<td>149</td>
<td>8.9</td>
<td>8.1</td>
<td>-0.8</td>
<td>-1.1 to -0.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Weight (pounds)</td>
<td>147</td>
<td>231.4</td>
<td>228.3</td>
<td>-3.0</td>
<td>-5.8 to -0.3</td>
<td>.03</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>147</td>
<td>105.0</td>
<td>103.6</td>
<td>-1.4</td>
<td>-2.6 to -0.1</td>
<td>.03</td>
</tr>
<tr>
<td>5% weight loss (%)</td>
<td>147</td>
<td>0.0</td>
<td>18.4</td>
<td>18.4</td>
<td>0.1 to 0.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Diabetes Distress Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Burden</td>
<td>167</td>
<td>2.6</td>
<td>2.3</td>
<td>-0.3</td>
<td>-0.5 to -0.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Physician-Related</td>
<td>167</td>
<td>2.7</td>
<td>2.4</td>
<td>-0.3</td>
<td>-0.5 to -0.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Regimen-Related</td>
<td>167</td>
<td>2.1</td>
<td>1.8</td>
<td>-0.3</td>
<td>-0.4 to -0.1</td>
<td>.001</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>167</td>
<td>3.0</td>
<td>2.6</td>
<td>-0.4</td>
<td>-0.6 to -0.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Medication adherence (%)</td>
<td>158</td>
<td>20.3</td>
<td>31.0</td>
<td>10.7</td>
<td>_c</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Elevated risk subsample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC (mg/dL)</td>
<td>43</td>
<td>230.0</td>
<td>190.5</td>
<td>-39.5</td>
<td>-51.3 to -27.6</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>114</td>
<td>131.6</td>
<td>132.5</td>
<td>0.9</td>
<td>-2.1 to 3.9</td>
<td>.54</td>
</tr>
<tr>
<td>DBP (mmHg)</td>
<td>114</td>
<td>84.7</td>
<td>82.0</td>
<td>-2.7</td>
<td>-4.3 to -1.0</td>
<td>.002</td>
</tr>
</tbody>
</table>

aStudy participants with complete data from both baseline and 4-month time points.
bHbA\textsubscript{1c}: hemoglobin A\textsubscript{1c}.
c—:Not applicable.
dStudy participants who began the study with elevated cardiovascular risk factors.
eTC: total cholesterol.
fSBP: systolic blood pressure.
gDBP: diastolic blood pressure.

Table 3. Baseline to post-test changes in hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) based on starting HbA\textsubscript{1c}.

<table>
<thead>
<tr>
<th>HbA\textsubscript{1c} category</th>
<th>n</th>
<th>Baseline</th>
<th>Post-test</th>
<th>Difference</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7.5%</td>
<td>24</td>
<td>6.3</td>
<td>6.4</td>
<td>0.1</td>
<td>-0.2 to 0.4</td>
<td>.49</td>
</tr>
<tr>
<td>7.5%–7.9%</td>
<td>24</td>
<td>7.7</td>
<td>7.4</td>
<td>-0.3</td>
<td>-0.6 to 0.1</td>
<td>.18</td>
</tr>
<tr>
<td>8.0%–8.9%</td>
<td>28</td>
<td>8.4</td>
<td>7.8</td>
<td>-0.6</td>
<td>-1.0 to -0.2</td>
<td>.002</td>
</tr>
<tr>
<td>&gt;9.0%</td>
<td>73</td>
<td>10.4</td>
<td>9.0</td>
<td>-1.4</td>
<td>-1.8 to -0.9</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Cardiovascular Outcomes
At baseline, 58.5% (114/195) of the participants had systolic or diastolic blood pressure above the normal range (<120 mmHg and <80 mmHg, respectively). There was no significant change in systolic blood pressure, whereas diastolic blood pressure decreased by an average of 2.7 mmHg ($t_{113}=-3.2$, $P=.002$). Only 43 participants had elevated TC above 200 mg/dL at baseline, and a significant decrease was found post-test ($t_{42}=-6.7$, $P<.001$) (Table 2).

Patient-Reported Outcomes
In the total sample, diabetes distress significantly decreased from 2.6 at baseline to 2.3 at post-test ($t_{166}=-4.5$, $P<.001$; Table 2). Significant improvements in distress were observed across all DDS subscales ($P<.01$). The proportion of the sample adherent to their medication regimen increased from 20% at baseline to 31% at post-test (McNemar $\chi^2_{1,158}=7.0$, $P=.01$).

Discussion
Principal Findings
The results of this study provide initial evidence that the enhanced digital DSMES program was effective for improving HbAlc, weight, diabetes distress, and medication adherence among a sample of people with T2DM and elevated HbA\textsubscript{1c}. Furthermore, those who were furthest from their HbA\textsubscript{1c} goal at the start of the program (baseline HbA\textsubscript{1c} $\geq$ 9.0%) achieved the greatest improvement in HbA\textsubscript{1c}, with an average change of 1.4%. 

https://diabetes.jmir.org/2021/1/e25295 JMIR Diabetes 2021 | vol. 6 | iss. 1 | e25295 | p.92
We found an inconsistent impact on cardiovascular outcomes among participants who started the study with elevated risk factors, with some improvements in diastolic blood pressure and TC, but no improvements in systolic blood pressure. However, blood pressure at baseline was close to the nationally recommended goal for those with diabetes, and the program was not designed to address hypertension specifically. Engagement was strong as evidenced by the high frequency of use across the features of the digital platform.

These results are consistent with prior studies of digital DSMES programs (both academic and commercial) that showed improvements in HbA1c and psychosocial outcomes [3,25-28]. In particular, the magnitude of the HbA1c reduction in this program is comparable to that of prior studies. Kumar et al [15] reported an HbA1c reduction of 0.86% and a higher effect in those with a higher baseline HbA1c. Dixon et al [16] reported a higher reduction in HbA1c by baseline group, but the intervention also included medication titration and physician support. This study adds to the growing evidence that digital DSMES significantly improves HbA1c and can also impact weight loss and cholesterol [12,29].

The clinical outcomes observed in this study meet or exceed those expected from traditional DSMES programs as set by the American Diabetes Association [30], as well as more resource-intensive digitally delivered programs that combine DSMES with physician telehealth services [16]. Further, the high rates of participant engagement with the program highlight many of the benefits of continuously accessible DSMES.

The improvements in medication adherence are encouraging given that this is a major challenge in diabetes management [31-33]. Digital delivery offers unique opportunities for patient engagement around improving medication-taking behaviors, as CDCES staff can be more proactive and support medication use in a timelier manner. Mobile apps can surface more frequent screenings, follow up, and in-app tracking to identify issues sooner so that a CDCES can reach out and provide education and support.

Limitations
There were several limitations to this pilot study. First, this pilot study is limited by its single-arm design and therefore carries the typical challenges in a nonrandomized design of unknown causal inference. Future research will benefit from a control group comparison and a randomized design to allow for a maximally rigorous test of the intervention. Second, we had to change the study methodology for follow-up lab measurement due to COVID-19 by shifting to a self-collected blood specimen versus a phlebotomist-collected venipuncture specimen; this creates potential for measurement error between instruments. However, this risk is attenuated by the high correlation of the venipuncture HbA1c and dried blood spot method [21]. Third, it is possible that the study sample recruited may not be fully representative or generalizable of the population of people living with diabetes, as participants self-selected from the online health community into the research opportunity. However, the clinical criteria (ie, HbA1c outside of the desired therapeutic range) increases the likelihood that study participants were individuals who would benefit from better diabetes self-management. Despite the high rates of program engagement observed among participants across the 4-month study, expectations around engagement in digital health studies remain exploratory, with varying definitions of meaningful engagement across digital platforms.

Conclusions
This study provides additional evidence that a digitally delivered DSMES program enhanced with deep lifestyle and behavior change support impacts HbA1c for people with T2DM and elevated HbA1c, showing the greatest benefit for those with higher blood glucose levels, and suggests benefits for weight loss and improvements in cardiovascular outcomes. Future research is needed to understand the potential impact of digital DSMES on long-term diabetes outcomes to meet the needs of the changing health care landscape.

Acknowledgments
The authors would like to thank Andrea Newcom, Bailey Peterka, Carolyn Salter, Danene Moberly, Melinda Merry, and Brieana Polk-Perez for their support of the project and work with participants. We would also like to thank Sara Cross and Anna Telthorst from Quest Diagnostics, and Kimberly Russell, Lisa Johnstone, Amber Hogue, and Maximo Prescott from Evidation Health for study management. Data included in this manuscript were presented in an abstract at the 20th Annual Diabetes Technology Meeting Virtual Poster Session on November 19, 2020. This study was funded by Omada Health, Inc.

Conflicts of Interest
FWA, RQ, CCS, MT, and CBJ are employees of Omada Health, Inc, and receive salary and stock options. CC and JJ are employees of Evidation Health, Inc, and receive salary. Evidation Health, Inc received funds from Omada Health, Inc to perform the study.

References


Abbreviations

CDCES: certified diabetes care and education specialist
DDS: Diabetes Distress Scale
DSMES: diabetes self-management education and support
HbA1c: hemoglobin A1c
HDL: high-density lipoprotein
LDL: low-density lipoprotein
PSC: Patient Service Center
SMAQ: Simplified Medication Adherence Questionnaire
T2DM: type 2 diabetes mellitus
TC: total cholesterol

©Folasade Wilson-Anumudu, Ryan Quan, Cynthia Castro Sweet, Christian Cerrada, Jessie Juusola, Michael Turken, Carolyn Bradner Jasik. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 22.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Diabetes Engagement and Activation Platform for Implementation and Effectiveness of Automated Virtual Type 2 Diabetes Self-Management Education: Randomized Controlled Trial

Roy Sabo1, PhD; Jo Robins1, PhD; Stacy Lutz2, BSc; Paulette Kashiri1, MPH; Teresa Day1, MSc; Benjamin Webel1, BA; Alex Krist1, MD

1Virginia Commonwealth University, Richmond, VA, United States
2Privia Health, LLC, Arlington, VA, United States

Corresponding Author:
Roy Sabo, PhD
Virginia Commonwealth University
830 East Main Street
Richmond, VA
United States
Phone: 1 804 828 3047
Email: roy.sabo@vcuhealth.org

Abstract

Background: Patients with type 2 diabetes require recommendations for self-management education and support.

Objective: In this study, we aim to design the Diabetes Engagement and Activation Platform (DEAP)—an automated patient education tool integrated into primary care workflow—and examine its implementation and effectiveness.

Methods: We invited patients aged 18-85 years with a hemoglobin A1c (HbA1c) level ≥8 to participate in a randomized controlled trial comparing DEAP with usual care. DEAP modules addressing type 2 diabetes self-management education and support domains were programmed into patient portals, each with self-guided educational readings, videos, and questions. Care teams received patient summaries and were alerted to patients with low confidence or requesting additional help. HbA1c, BMI, and systolic and diastolic blood pressure (DBP) were measured.

Results: Out of the 680 patients invited to participate, 337 (49.5%) agreed and were randomized. All of the 189 intervention patients accessed the first module, and 140 patients (74.1%) accessed all 9 modules. Postmodule knowledge and confidence scores were high. Only 18 patients requested additional help from the care team. BMI was lower for intervention patients than controls at 3 months (31.7 kg/m² vs 32.1 kg/m²; P=.04) and 6 months (32.5 kg/m² vs 33.0 kg/m²; P=.003); improvements were even greater for intervention patients completing at least one module. There were no differences in 3- or 6-month HbA1c or blood pressure levels in the intent-to-treat analysis. However, intervention patients completing at least one module compared with controls had a better HbA1c level (7.6% vs 8.2%; P=.03) and DBP (72.3 mm Hg vs 75.9 mm Hg; P=.01) at 3 months.

Conclusions: The findings of this study concluded that a significant proportion of patients will participate in an automated virtual diabetes self-management program embedded into patient portals and health systems show promise in helping patients manage their diabetes, weight, and blood pressure.

Trial Registration: ClinicalTrials.gov NCT02957721; https://clinicaltrials.gov/ct2/show/NCT02957721

(JMIR Diabetes 2021;6(1):e26621) doi:10.2196/26621

KEYWORDS

type 2 diabetes mellitus; self-management education; patient engagement; informatics
**Introduction**

**Background**

Type 2 diabetes (T2D) affects an estimated 34 million people in the United States [1], costing US $327 billion annually [2]. T2D prevalence in the United States is expected to increase, whereas costs are expected to double over the next 25 years [3,4]. T2D self-management education and support (DSMES) provides individuals with the information and problem-solving skills needed to self-manage T2D and has been shown to improve medication adherence, self-blood glucose monitoring, glycemic control, and dietary behaviors [5,6] and reduce complications from uncontrolled T2D [7,8]. The American Diabetes Association (ADA) recommends the provision of DSMES for every patient at 4 points: at diagnosis, annually thereafter, when complicating factors arise, and when transitioning to new care teams [9].

Despite its proven effectiveness, many patients do not receive DSMES. Of the patients referred, only 23%-66% follow through to receive DSMES [10] because of barriers such as time commitments, schedule conflicts, or transportation difficulties [7]. Innovative DSMES delivery methods are needed to better meet patients’ needs and leverage limited resources.

Health information technology, specifically personal health records (PHRs) integrated into electronic health records (EHRs), has the potential to increase patient access to DSMES by automating the provision of educational content and allowing patients to review and complete programs at convenient times and locations [11]. Integrated PHRs can help automate identifying patients needing additional help, allow patients to initiate requests for support, and alert team members to initiate care or direct patients to existing community resources [12,13].

**Objectives**

To help leverage the benefits of health information technology in providing DSMES, we created the Diabetes Engagement and Activation Platform (DEAP), which is an automated patient educational tool integrated directly into the primary care workflow. DEAP is accessed from the patient portal, consists of 9 modules that address the recommended ADA domains of diabetes education, assesses patients’ knowledge and confidence in managing each domain, and alerts care team members of patient needs. We aim to conduct a randomized controlled trial (RCT) to evaluate the implementation of DEAP and its effectiveness relative to usual care for improving patient T2D outcomes.

**Methods**

**Overview**

We conducted a patient-level RCT evaluating the implementation and effectiveness of DEAP with respect to changes in glycated hemoglobin (HbA1c; primary outcome), BMI, and blood pressure (BP) from baseline to 3 and 6 months. The study was conducted between November 1, 2017, and May 7, 2018, to achieve 6 months of patient tracking. This study was approved by the Virginia Commonwealth University Institutional Review Board and registered at ClinicalTrials.gov (identifier NCT02957721).

**Setting**

A total of 21 practices spanning 5 states from the Privia Health, LLC (Privia), a technology-enabled, physician enablement company that collaborates with medical groups, health plans, and health systems, were recruited to participate in this study. The practices predominantly serve commercially insured populations and those covered by Medicare.

**Patient Sampling**

All patients aged between 18 and 85 years with a T2D diagnosis, HbA1c ≥8.0%, and practice portal account were sent an email to participate by their primary care clinician. Identification was automated in the practices’ EHR, and the email was sent 2 days after a laboratory result with an elevated HbA1c level. The automated email, addressed by the primary care clinician, asked the patient to log in to the portal, which alerted the patient that their diabetes seemed poorly controlled. The system randomized patients in a 1:1 manner to receive either DEAP (intervention) or 1 page of information about diabetes (usual care control). No blinding or allocation concealment was used in this study.

**Intervention and Control Conditions**

DEAP was integrated into the practices’ EHR, patient portal, and data warehouse. DEAP consisted of 9 self-directed DSMES modules for patients and care team alerts for clinicians to assist patients requesting additional help. The DEAP modules covered the Standard 6: Curriculum from the National Standards for Diabetes Self-Management Education and Support [14]. The 9 modules included: (1) diabetes disease process and general treatment, (2) nutritional management, (3) physical activity, (4) medications, (5) monitoring blood glucose, (6) acute complications, (7) chronic complications, (8) mental health, and (9) goal setting. Patients were sent modules in order and received biweekly reminders until they completed the modules. The next module was sent when a patient completed a module or after 7 days of noncompletion, which allowed patients to skip or ignore the modules.

Each module included 1 to 3 handouts and 1 to 3 videos for patients to review (Multimedia Appendix 1). Content was selected from existing publicly available and validated material from the ADA, National Diabetes Education Program, American Association of Diabetes Educators, Mayo Clinic, MedlinePlus, and other sources. Content was selected by the research team with support from 2 certified diabetic educators, a lay community educator, and 2 patients with T2D. Inclusion criteria for content consisted of being clear and understandable, evidence based, and engaging. Upon completion of a module, patients were asked 4 questions to assess their knowledge, 1 question to assess their confidence in managing the module’s domain, and 1 question to understand if the patient wanted additional help from the care team related to the content in the module. DEAP sent a summary of the patient’s responses to the primary clinician and provided an alert for patients reporting low confidence or requesting help in managing a domain.
Patients randomized to the usual care control group received 1 page of general diabetes information, which was equivalent to the handout information in the first DEAP module. They did not have access to the structured DEAP curriculum, knowledge or confidence assessments, or care team alerts.

**Measurements and Informatics**

The patient portal and Privia electronic data warehouse were used to track patient progress through the curriculum, indicate whether modules were accessed and completed (completion was measured as a patient answering all postmodule questions), and record responses to end-module questions. The EHR was used to determine patient eligibility, measure patient characteristics (gender, age, race, ethnicity, preferred language, and insurance type), and capture health outcomes (HbA\(_1c\), BMI, and BP). Health outcomes for measuring effectiveness included HbA\(_1c\) (primary outcome) and BMI and BP (secondary outcomes), captured at baseline, 3 months, and 6 months. Implementation measures consisted of knowledge, confidence, adoption, and reach. Confidence was assessed using a Likert scale ranging from not confident at all to completely confident. Adoption was defined as the number of practices that were willing to participate in the study. We defined reach as the percentage of patients who agreed to participate in the study, the percentage of patients who started the DEAP curriculum within the intervention group, the percentage of patients who completed the DEAP curriculum, and the total number of DEAP modules that were accessed.

**Statistical Analysis and Sample Size Justification**

We conducted both an intent-to-treat analysis of all intervention versus usual care control patients and a per-protocol analysis of intervention patients who completed at least one module (representing minimal intervention exposure) versus control patients. For both models, we made baseline-adjusted comparisons of 3- and 6-month means for HbA\(_1c\), BMI, and systolic BP (SBP) and diastolic BP (DBP) between the study groups. Using linear mixed models, health outcomes (HbA\(_1c\), BMI, and BP) at 3 and 6 months were modeled against a 2-level fixed group effect (intervention or control), the baseline value of that health outcome measurement, and a group-baseline interaction effect; the interaction term was removed if it was not significant at the 10% level and the Bayesian Information Criterion was lower in the no-interaction model. As an additional sensitivity analysis, unadjusted comparisons of the change in mean HbA\(_1c\), BMI, and BP over time and between the study groups were made using linear mixed models, including continuous health outcomes (HbA\(_1c\), BMI, and BP), a 2-level fixed group effect (intervention or control), a 3-level fixed time effect (baseline, 3 months, and 6 months), a fixed group-time interaction effect, and a patient-level random effect to account for within-participant dependence because of repeated measurements over time. The MEANS, FREQ, and GLIMMIX procedures in SAS statistical software (version 9.4) were used for analysis.

Sample size calculations were based on the assumption that 50% of participants would either decline to participate or not complete the study; therefore, recruiting 320 eligible participants would help ensure that 80 patients would participate and finish the study in each group (160 in total). Assuming a 5% type I error rate and an HbA\(_1c\) SD of 2\([4,15]\), we estimated over 80% power to declare mean HbA\(_1c\) for the intervention group to be significantly lower than in the usual care control group at either 3 or 6 months by at least 1 unit.

**Results**

**Implementation Analyses**

**Adoption**

The original plan was to recruit 4 practices from Privia’s network. However, we encountered significant practice enthusiasm across the organization, and a total of 21 practices across 5 states participated in the study. After the study was completed, Privia’s network extended DEAP to all practices as part of their standard operations.

**Reach**

The frequencies and percentages of intervention patients who accessed each of the training modules (and the numbers and percentages of those patients answering at least one question in each module and completing each module) are reported in Tables 1 and 2. Of the 189 intervention patients accessing at least the first module, the vast majority (140/189, 74.1%) eventually accessed all 9 modules, whereas only a few (8/189, 4.2%) failed to continue. Between 14% (21/151) and 28% (54/189) of the patients starting each module answered at least one of the corresponding postmodule questions. Of the 63 patients who answered at least one question in any module, 53 (84%) completed the questions to at least one module, with the majority answering at least one question completing all questions in each module.
### Table 1. Intervention patients (n=189) who accessed, started, and completed particular Diabetes Engagement and Activation Platform modules.

<table>
<thead>
<tr>
<th>Module</th>
<th>Access (n=189), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Started&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Completed&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total participants, n</td>
<td>n (%)</td>
<td>Total participants, n</td>
</tr>
<tr>
<td>1. Basic assessment</td>
<td>189 (100.0)</td>
<td>189</td>
<td>54 (28.6)</td>
</tr>
<tr>
<td>2. Nutrition</td>
<td>181 (95.8)</td>
<td>181</td>
<td>34 (18.7)</td>
</tr>
<tr>
<td>3. Exercise</td>
<td>173 (91.5)</td>
<td>173</td>
<td>36 (20.8)</td>
</tr>
<tr>
<td>4. Mediations</td>
<td>167 (88.4)</td>
<td>167</td>
<td>25 (15.0)</td>
</tr>
<tr>
<td>5. Blood sugar</td>
<td>160 (84.6)</td>
<td>160</td>
<td>25 (15.6)</td>
</tr>
<tr>
<td>6. Acute complications</td>
<td>154 (81.4)</td>
<td>154</td>
<td>25 (16.2)</td>
</tr>
<tr>
<td>7. Chronic diabetes</td>
<td>151 (79.8)</td>
<td>151</td>
<td>21 (13.9)</td>
</tr>
<tr>
<td>8. Mood</td>
<td>146 (77.2)</td>
<td>146</td>
<td>22 (15.1)</td>
</tr>
<tr>
<td>9. Healthy goals</td>
<td>140 (74.1)</td>
<td>140</td>
<td>20 (14.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentage calculated as 100 × (frequency accessed/189)%.
<sup>b</sup>Percentage calculated as 100 × (frequency started/frequency accessed)%.
<sup>c</sup>Percentage calculated as 100 × (frequency completed/frequency started)%.

### Table 2. Number of Diabetes Engagement and Activation Platform modules accessed, started, and completed by intervention patients (n=189).

<table>
<thead>
<tr>
<th>Number of modules accessed, n</th>
<th>Accessed, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Started, n (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Completed, n (%)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
<td>126 (66.6)</td>
<td>136 (71.9)</td>
</tr>
<tr>
<td>1</td>
<td>8 (4.2)</td>
<td>24 (12.6)</td>
<td>16 (8.4)</td>
</tr>
<tr>
<td>2</td>
<td>8 (4.2)</td>
<td>5 (2.6)</td>
<td>7 (3.7)</td>
</tr>
<tr>
<td>3</td>
<td>6 (3.1)</td>
<td>7 (3.7)</td>
<td>6 (3.1)</td>
</tr>
<tr>
<td>4</td>
<td>7 (3.7)</td>
<td>3 (1.5)</td>
<td>2 (1.0)</td>
</tr>
<tr>
<td>5</td>
<td>6 (3.1)</td>
<td>2 (1.0)</td>
<td>2 (1.0)</td>
</tr>
<tr>
<td>6</td>
<td>3 (1.5)</td>
<td>1 (0)</td>
<td>4 (2.1)</td>
</tr>
<tr>
<td>7</td>
<td>5 (2.6)</td>
<td>4 (2.1)</td>
<td>2 (1.0)</td>
</tr>
<tr>
<td>8</td>
<td>6 (3.1)</td>
<td>2 (1.0)</td>
<td>9 (4.7)</td>
</tr>
<tr>
<td>9</td>
<td>140 (74.0)</td>
<td>15 (7.9)</td>
<td>5 (2.6)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentage calculated as 100 × (frequency accessed/189)%; mean 7.7, SD 2.5.
<sup>b</sup>Percentage calculated as 100 × (frequency started/189)%; mean 1.4, SD 2.7.
<sup>c</sup>Percentage calculated as 100 × (frequency completed/189)%; mean 1.2, SD 2.5.
<sup>d</sup>N/A: not applicable.

**Patient Knowledge, Confidence, and Help Seeking**

Patients answered a majority of knowledge questions correctly for each module (Table 3). The 4 most commonly missed questions included understanding what the HbA1c measured, causes of low blood sugar, recommended number of daily servings of fruits and vegetables, and strategies for reducing cardiovascular risk. Upon completion of a module, most patients reported being very or completely confident of the module’s content. Only 18 patients asked for additional help from the care team after completing a module, most commonly after completing the introduction module (9/54, 17%), nutrition module (4/33, 12%), and exercise module (2/35, 6%).
Table 3. Summaries of knowledge assessment, confidence question, and desire to be contacted for each Diabetes Engagement and Activation Platform module.

<table>
<thead>
<tr>
<th>Module</th>
<th>Correct knowledge questions</th>
<th>Confidence question</th>
<th>Expressed desire to be contacted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size, n</td>
<td>Mean (SD)</td>
<td>Not or a little confident, n (%)</td>
</tr>
<tr>
<td>1. Basic assessment</td>
<td>34</td>
<td>3.6 (0.54)</td>
<td>12 (24)</td>
</tr>
<tr>
<td>2. Nutrition</td>
<td>33</td>
<td>2.9 (0.77)</td>
<td>16 (47)</td>
</tr>
<tr>
<td>3. Exercise</td>
<td>32</td>
<td>3.7 (0.52)</td>
<td>16 (47)</td>
</tr>
<tr>
<td>4. Mediations</td>
<td>23</td>
<td>3.7 (0.54)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>5. Blood sugar</td>
<td>23</td>
<td>3.7 (0.65)</td>
<td>8 (33)</td>
</tr>
<tr>
<td>6. Acute complications</td>
<td>23</td>
<td>3.3 (0.88)</td>
<td>7 (28)</td>
</tr>
<tr>
<td>7. Chronic complications</td>
<td>21</td>
<td>3.0 (0.38)</td>
<td>6 (29)</td>
</tr>
<tr>
<td>8. Mood</td>
<td>17</td>
<td>3.7 (0.77)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>9. Healthy goals</td>
<td>15</td>
<td>3.9 (0.26)</td>
<td>4 (21)</td>
</tr>
<tr>
<td>All modules</td>
<td>5</td>
<td>31.8 (2.17)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Sample sizes for each column can be different.

bN/A: not applicable.

**Effectiveness Analyses**

A total of 680 patients met the eligibility criteria and were emailed the portal invitation (Figure 1). Of those, 343 either never opened the portal message or after opening the message decided not to proceed with participation. Of the remaining 337 patients, 189 were randomly allocated to the intervention group and 148 to the control group. We identified 327 of the allocated patients in the EHR group (183 patients in the intervention group and 144 patients in the control group). All intervention patients (100%) accessed the first training module, with a percentage decrease for each successive module, and 74% (140/189) accessed the ninth module. Between 14% (21/151) and 28% (54/189) of the patients accessing the modules answered at least one of the corresponding postmodule questions, and 53 completed at least one module. A summary of patient characteristics and demographics are presented in Table 4. The average patient was just above 60 years, had an HbA1c level >9, had a BMI in the obese range (>30), and had controlled BP (SBP<140). Both groups had similar rates of men and women, whereas the majority of participants were non-Hispanic, White, with English as their preferred language. Most participants had commercial health insurance or Medicare.
Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.
Table 4. Patient demographics at baseline.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Total participants, n</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>183</td>
<td>144</td>
</tr>
<tr>
<td>HbA1c&lt;sup&gt;a&lt;/sup&gt;, mean (SD)</td>
<td>180</td>
<td>142</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), mean (SD)</td>
<td>179</td>
<td>136</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg), mean (SD)</td>
<td>180</td>
<td>137</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg), mean (SD)</td>
<td>180</td>
<td>136</td>
</tr>
<tr>
<td>Sex, n (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>Male</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>155</td>
<td>113</td>
</tr>
<tr>
<td>Black</td>
<td>155</td>
<td>113</td>
</tr>
<tr>
<td>Other</td>
<td>155</td>
<td>113</td>
</tr>
<tr>
<td>White</td>
<td>155</td>
<td>113</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>137</td>
<td>97</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>137</td>
<td>134</td>
</tr>
<tr>
<td>Language, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-English</td>
<td>176</td>
<td>136</td>
</tr>
<tr>
<td>English</td>
<td>176</td>
<td>136</td>
</tr>
<tr>
<td>Insurance type, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>Medicaid</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>None</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>Commercial</td>
<td>183</td>
<td>143</td>
</tr>
<tr>
<td>Unknown</td>
<td>183</td>
<td>143</td>
</tr>
</tbody>
</table>

<sup>a</sup>HbA1c: glycated hemoglobin.

<sup>b</sup>Percentage of sample with an event.

**Intent-to-Treat Analysis**

Table 5 contains summaries of the comparisons of mean health outcomes between intervention and control groups. There was no evidence that the mean for the primary outcome (HbA1c) was lower in the intervention group than in the control group at 3 months (8.0% vs 8.2%; \( P=.38 \)) or at 6 months (8.2% vs 8.4%; \( P=.27 \)). The mean BMI was significantly reduced in intervention group patients relative to control group patients at 3 months (31.7 kg/m\(^2\) vs 32.1 kg/m\(^2\); \( P=.04 \)) and at 6 months (32.5 kg/m\(^2\) vs 33.0 kg/m\(^2\); \( P=.02 \)). There was no evidence of improved SBP or DBP in the intervention group patients compared with the controls. Results were similar in the changes comparison analyses (Table 6), with no evidence of differences in baseline and 3-month changes between groups for any measures, and with only the change in BMI between baseline and 6 months for intervention group patients (–0.4 kg/m\(^2\) decrease vs 0.1 kg/m\(^2\) increase; \( P=.02 \)).
Table 5. Comparisons of baseline-adjusted health outcome means between groups at 3 and 6 months.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Intervention</th>
<th></th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completed ≥1 module</td>
<td>All intervention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>Mean (95% CI)</td>
<td>n</td>
</tr>
<tr>
<td>HbA1c</td>
<td>3 months</td>
<td>36</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>6 months</td>
<td>25</td>
<td>95</td>
</tr>
<tr>
<td>BMI</td>
<td>3 months</td>
<td>40</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>6 months</td>
<td>33</td>
<td>120</td>
</tr>
<tr>
<td>SBP</td>
<td>3 months</td>
<td>40</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>6 months</td>
<td>32</td>
<td>122</td>
</tr>
<tr>
<td>DBP</td>
<td>3 months</td>
<td>40</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>6 months</td>
<td>32</td>
<td>122</td>
</tr>
</tbody>
</table>

*Mean: baseline-adjusted sample predicted value.

HbA1c: glycated hemoglobin.

†Intent-to-treat (ITT) analysis (comparison between intervention and control patients; control-intervention): difference=0.2, 95% CI –0.2 to 0.6; P=.38 (indicates the interaction term left in the model). Per-protocol (PP) analysis: comparison between intervention subjects completing at least one Diabetes Engagement and Activation Platform module (answering postmodule questions) and control patients. PP analysis (control-intervention): difference=0.6, 95% CI 0.1 to 1.1; P=.03.

‡ITT analysis (control-intervention): difference=0.3, 95% CI –0.2 to 0.8; P=.27. PP analysis (control-intervention): difference=0.5, 95% CI –0.2 to 1.2; P=.17.

§ITT analysis (control-intervention): difference=0.4, 95% CI 0.0 to 0.8; P=.04 (indicates the interaction term left in the model). PP analysis (control-intervention): difference=1.0, 95% CI 0.5 to 1.4; P<.001.

‖ITT analysis (control-intervention): difference=0.5, 95% CI 0.1 to 1.0; P=.02. PP analysis (control-intervention): difference=1.0, 95% CI 0.5 to 1.5; P<.001.

¶SBP: systolic blood pressure.

‖ITT analysis (control-intervention): difference=0.7, 95% CI –3.4 to 4.9; P=.73. PP analysis (control-intervention): difference=3.2, 95% CI –2.3 to 8.8; P=.25.

‖ITT analysis (control-intervention): difference=0.2, 95% CI –4.0 to 4.3; P=.94. PP analysis (control-intervention): difference=0.5, 95% CI –4.9 to 5.9; P=.85.

‡DBP: diastolic blood pressure.

‖ITT analysis (control-intervention): difference=1.1, 95% CI –1.4 to 3.6; P=.39. PP analysis (control-intervention): difference=4.3, 95% CI 1.0 to 7.5; P=.01.

‡ITT analysis (control-intervention): difference=0.4, 95% CI –2.5 to 3.4; P=.78. PP analysis (control-intervention): difference=1.6, 95% CI –1.9 to 5.1; P=.37.
### Table 6. Comparison between groups of change in glycated hemoglobin, BMI, and blood pressure from baseline to 3 and 6 months.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Intervention</th>
<th>Completed ≥1 module</th>
<th>All interventions</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>Mean (95% CI)</td>
<td>n</td>
</tr>
<tr>
<td>HbA1c b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 3 months c</td>
<td></td>
<td>36</td>
<td>−1.8 (−2.4 to −1.3)</td>
<td>106</td>
</tr>
<tr>
<td>Baseline to 6 months d</td>
<td></td>
<td>25</td>
<td>−1.5 (−2.2 to −0.8)</td>
<td>95</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 3 months c</td>
<td></td>
<td>40</td>
<td>−0.9 (−1.3 to −0.6)</td>
<td>138</td>
</tr>
<tr>
<td>Baseline to 6 months f</td>
<td></td>
<td>33</td>
<td>−0.8 (−1.3 to −0.4)</td>
<td>119</td>
</tr>
<tr>
<td>SBP g</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 3 months b</td>
<td></td>
<td>40</td>
<td>−5.0 (−10.2 to 0.2)</td>
<td>135</td>
</tr>
<tr>
<td>Baseline to 6 months i</td>
<td></td>
<td>32</td>
<td>−1.7 (−6.7 to 3.4)</td>
<td>120</td>
</tr>
<tr>
<td>DBP i</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 3 months k</td>
<td></td>
<td>40</td>
<td>−5.2 (−8.1 to −2.2)</td>
<td>135</td>
</tr>
<tr>
<td>Baseline to 6 months l</td>
<td></td>
<td>32</td>
<td>−2.6 (−5.9 to 0.8)</td>
<td>120</td>
</tr>
</tbody>
</table>

\( ^a \)Mean is the model-predicted difference (baseline minus the 3- or 6-month value).

\( ^b \)HbA1c: glycated hemoglobin.

\( ^c \)Intent-to-treat (ITT) analysis (control-intervention): difference=−0.2, 95% CI −0.6 to 0.3; \( P=0.53 \) (comparison between all intervention and control patients). Per-protocol (PP) analysis (control-intervention): difference=0.3, 95% CI −0.3 to 1.0; \( P=0.29 \) (comparison between intervention subjects completing at least one DEAP module [answering postmodule questions] and control patients).

\( ^d \)ITT analysis (control-intervention): difference=−0.1, 95% CI −0.7 to 0.4; \( P=0.67 \). PP analysis (control-intervention): difference=0.2, 95% CI −0.6 to 1.1; \( P=0.54 \).

\( ^e \)ITT analysis (control-intervention): difference=0.3, 95% CI 0.0 to 0.7; \( P=0.07 \). PP analysis (control-intervention): difference=1.0, 95% CI 0.5 to 1.4; \( P<0.001 \).

\( ^f \)ITT analysis (control-intervention): difference=0.5, 95% CI 0.1 to 0.9; \( P=0.02 \). PP analysis (control-intervention): difference=1.0, 95% CI 0.5 to 1.5; \( P<0.001 \).

\( ^g \)SBP: systolic blood pressure.

\( ^h \)ITT analysis (control-intervention): difference=2.1, 95% CI −1.9 to 6.2; \( P=0.30 \). PP analysis (control-intervention): difference=3.3, 95% CI −2.8 to 9.4; \( P=0.28 \).

\( ^i \)ITT analysis (control-intervention): difference=−0.7, 95% CI −4.6 to 3.2; \( P=0.73 \). PP analysis (control-intervention): difference=0.6, 95% CI −5.3 to 6.5; \( P=0.85 \).

\( ^j \)DBP: diastolic blood pressure.

\( ^k \)ITT analysis (control-intervention): difference=1.1, 95% CI −1.3 to 3.6; \( P=0.35 \). PP analysis (control-intervention): difference=3.9, 95% CI 0.4 to 7.4; \( P=0.03 \).

\( ^l \)ITT analysis (control-intervention): difference=−1.0, 95% CI −3.8 to 1.8; \( P=0.47 \). PP analysis (control-intervention): difference=1.3, 95% CI −2.7 to 5.3; \( P=0.52 \).

**Per-Protocol Analyses**

Comparisons among intervention group patients completing at least one DEAP module and controls are also provided in Table 5. Those who completed at least one module had a lower mean HbA1c at 3 months compared with controls (7.6% vs 8.2%; \( P=0.03 \)), whereas there was no significant difference at 6 months (7.9% vs 8.4%; \( P=0.17 \)). Completers had significantly lower mean BMI at 3 months than controls (31.3 kg/m\(^2\) vs 32.1 kg/m\(^2\); \( P<0.001 \)) and at 6 months (31.6 kg/m\(^2\) vs 33.0 kg/m\(^2\); \( P<0.001 \)). There were no differences in SBP between completers and controls at 3 months (\( P=0.25 \)) and 6 months (\( P=0.85 \)). The intervention patients completing at least one module also had a larger mean DBP at 3 months than controls (72.3 mm Hg vs 75.9 mm Hg; \( P=0.01 \)), although there was no significant difference at 6 months (\( P=0.37 \)). Results from the comparison of change analyses (Table 6) were nearly identical, with the exception being that there was no evidence of different changes between groups in HbA1c at 3 months (\( P=0.29 \)) or 6 months (\( P=0.54 \)). The change in BMI was significantly larger in those who completed at least one module compared with controls between baseline and 3 months (−0.9 kg/m\(^2\) vs 0.1 kg/m\(^2\); \( P<0.01 \)) and 6 months (−0.8 kg/m\(^2\) vs 0.1 kg/m\(^2\); \( P<0.01 \)), and with the change in DBP significantly larger in those intervention group...
Discussion

Principal Findings

DEAP uses publicly available material in a systematic manner to automatically provide virtual diabetes education and support through pre-existing patient portals. DEAP Adoption exceeded what was expected to meet the study objectives, indicating that clinicians recognize the need for innovative, structured, accessible DSMES to optimize patient care and outcomes. With regard to reach, more patients accessed and used DEAP modules (74%) and then would access other simple educational messages sent to patients (about 20% of general Privia educational messages were opened by patients). This uptake of the automated DEAP content is similar to that of traditional in person DSMES classes [16]. DEAP facilitated high levels of confidence, knowledge, and help-seeking behaviors.

Although knowledge does not always correlate with improved self-management [17], the DEAP intervention group demonstrated improved BMI relative to controls, whereas our per-protocol analysis also showed evidence of improvement in HbA1c and DBP at 3 months postintervention for those completing modules. The lack of change in HbA1c and BP may be because of dilution from non-DEAP users, who did not change. Nonetheless, the improved BMI in the intent-to-treat analysis is particularly impressive, given that most interventions to help patients lose weight must be fairly intensive, often including 25 or more hours of contact over 6 months [18].

DEAP leverages the existing use of patient portals [19] and compiles existing patient educational materials and videos into an easily accessible and understandable format. A key element of DEAP’s success is the automatic identification of patients with elevated HbA1c within 2 days of the abnormal result, which removes the burden of identifying and engaging patients from the clinician and engages patients when they may be more amenable to making self-management changes. Another key element is that DEAP assembles publicly available information into a defined curriculum, making the material more acceptable and accessible to patients. Integrating DEAP into the clinician’s portal also comes with the imprimatur and credibility of the patient’s personal clinician.

Although we did observe benefits in this study comparing, we suspect that the benefits could have been greater if the automated self-directed learning was better coupled with support from the care team. How clinicians and care team members addressed the alerts was left to their discretion. Future implementations of DEAP could focus on alerting specific care team members when patients completed modules that could contact patients and offer additional ancillary services. For example, DEAP could notify a nutritionist when a patient expressed low confidence in managing their diet or missed a knowledge question [20] or a pharmacist about their medication management [21].

Limitations

A limitation of this study is the short time frame, as 6 months of follow-up may not be enough for DSMES to lead to substantial and sustainable behavioral or health changes. However, the shorter time frame resulted in a greater improvement in BMI observed in the intervention group compared with the control and the improved HbA1c, BMI, and DBP observed among DEAP users compared with nonusers. The generalization of these results may be limited by the predominantly White, English-speaking, and non-Hispanic study sample, although the use of multiple practices and the focus on patients seen in primary care are strengths. Another factor limiting generalization was requiring a patient portal account for inclusion; investigations of approaches to encourage portal uptake or delivery of DEAP through other mechanisms are warranted.

Conclusions

This low-intensity intervention to provide virtual diabetes self-management education proved both feasible and effective. The model is scalable, builds on existing infrastructures in many practices and health systems, and can be extended to other settings or conditions. Studying how automated self-directed approaches could be better linked with alerting care team members for additional directed care could have even greater benefits.

Acknowledgments

This work was supported by the National Center for Advancing Translational Sciences (UL1TR002649).

Authors' Contributions

PK, AK, SL, JR, and RS helped in the design of this trial. TD, PK, SL, and BW conducted the data collection and management activities. RS conducted the statistical analyses. AK, JR, and RS wrote the manuscript’s main draft. All authors reviewed the manuscript and provided comments, changes, and feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Diabetes Engagement and Activation Platform diabetes self-management education and support curriculum with sample content.
References


Abbreviations

ADA: American Diabetes Association  
BP: blood pressure  
DBP: diastolic blood pressure  
DEAP: Diabetes Engagement and Activation Platform  
DSMES: diabetes self-management education and support  
EHR: electronic health record  
HbA1c: glycated hemoglobin  
PHR: personal health record  
RCT: randomized controlled trial  
SBP: systolic blood pressure  
T2D: type 2 diabetes

©Roy Sabo, Jo Robins, Stacy Lutz, Paulette Kashiri, Teresa Day, Benjamin Webel, Alex Krist. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 29.03.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original work on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Ability of Current Machine Learning Algorithms to Predict and Detect Hypoglycemia in Patients With Diabetes Mellitus: Meta-analysis

Satoru Kodama¹, MD, PhD; Kazuya Fujihara², MD, PhD; Haruka Shiozaki³, PhD; Chika Horikawa¹, RD, PhD, CDE; Mayuko Harada Yamada², MD; Takaaki Sato², MD, PhD; Yuta Yaguchi², MD; Masahiko Yamamoto², MD; Masaru Kitazawa³, MD; Midori Iwanaga², MD; Yasuhiro Matsubayashi², MD, PhD; Hirohito Sone², MD, PhD, FACP

¹Department of Prevention of Noncommunicable Diseases and Promotion of Health Checkup, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan
²Department of Hematology, Endocrinology and Metabolism, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan
³Department of Health and Nutrition, Faculty of Human Life Studies, University of Niigata Prefecture, Niigata, Japan

Corresponding Author:
Satoru Kodama, MD, PhD
Department of Prevention of Noncommunicable Diseases and Promotion of Health Checkup
Niigata University Graduate School of Medical and Dental Sciences
1-757, Asahimachi-dori, Cyuoh-ku
Niigata, 951-8510
Japan
Phone: 81 25 227 2117
Email: ybbkodama@gmail.com

Abstract

Background: Machine learning (ML) algorithms have been widely introduced to diabetes research including those for the identification of hypoglycemia.

Objective: The objective of this meta-analysis is to assess the current ability of ML algorithms to detect hypoglycemia (ie, alert to hypoglycemia coinciding with its symptoms) or predict hypoglycemia (ie, alert to hypoglycemia before its symptoms have occurred).

Methods: Electronic literature searches (from January 1, 1950, to September 14, 2020) were conducted using the Dialog platform that covers 96 databases of peer-reviewed literature. Included studies had to train the ML algorithm in order to build a model to detect or predict hypoglycemia and test its performance. The set of 2 × 2 data (ie, number of true positives, false positives, true negatives, and false negatives) was pooled with a hierarchical summary receiver operating characteristic model.

Results: A total of 33 studies (14 studies for detecting hypoglycemia and 19 studies for predicting hypoglycemia) were eligible. For detection of hypoglycemia, pooled estimates (95% CI) of sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) were 0.79 (0.75-0.83), 0.80 (0.64-0.91), 8.05 (4.79-13.51), and 0.18 (0.12-0.27), respectively. For prediction of hypoglycemia, pooled estimates (95% CI) were 0.80 (0.72-0.86) for sensitivity, 0.92 (0.87-0.96) for specificity, 10.42 (5.82-18.65) for PLR, and 0.22 (0.15-0.31) for NLR.

Conclusions: Current ML algorithms have insufficient ability to detect ongoing hypoglycemia and considerate ability to predict impending hypoglycemia in patients with diabetes mellitus using hypoglycemic drugs with regard to diagnostic tests in accordance with the Users’ Guide to Medical Literature (PLR should be ≥5 and NLR should be ≤0.2 for moderate reliability). However, it should be emphasized that the clinical applicability of these ML algorithms should be evaluated according to patients’ risk profiles such as for hypoglycemia and its associated complications (eg, arrhythmia, neuroglycopenia) as well as the average ability of the ML algorithms. Continued research is required to develop more accurate ML algorithms than those that currently exist and to enhance the feasibility of applying ML in clinical settings.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020163682; http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42020163682

(JMIR Diabetes 2021;6(1):e22458) doi:10.2196/22458
KEYWORDS
machine learning; hypoglycemia; meta-analysis

Introduction

Hypoglycemia is a major barrier to achieving the tight glycemic control in patients with diabetes mellitus (DM) that is required to delay the progression of late DM-related complications. Although many patients exhibit symptoms of hypoglycemia such as anxiety, heart palpitations, and confusion, a significant number have diminished ability to recognize these hypoglycemic symptoms [1,2], which is defined as “impaired awareness of hypoglycemia” [3]. This impaired awareness can lead to severe hypoglycemia, which is associated with seizures, coma, and death. Real-time glucose monitoring can help patients maintain optimal glycemic control while avoiding symptomatic or asymptomatic hypoglycemia [4]. However, the traditional monitoring method, intermittent glucose monitoring by finger stick, provides only a limited number of readings and is unlikely to detect hypoglycemia of a short duration. Continuous glucose monitoring (CGM) typically produces a reading every 5 minutes and can alert the patient to not only the occurrence of hypoglycemia but also impending hypoglycemia [5]. Accuracy of CGM has progressively improved, with overall measurement errors reduced by twofold than in the first commercially available CGM devices introduced in 2000 [5].

However, even if CGM advancements enabled patients to continuously track their subcutaneous glucose levels, the statistical disadvantage of the CGM data stream would remain as a major limitation. The autocorrelation of the CGM reading vanishes after 30 minutes, meaning that the projection of blood glucose levels more than 30 minutes ahead would be inaccurate [6]. This finding suggests that the algorithm for identifying hypoglycemia should consider a patient’s contextual information such as diet, physical activity, and medications (including insulin) as well as various features of the CGM trend arrow [7].

Machine learning (ML) algorithms have been widely introduced to diabetes research including those for identification of hypoglycemia. The growing use of mobile health (mHealth) apps, sensors, wearables, and other point-of-care devices, including CGM sensors for self-monitoring and management of DM, have made possible the generation of automated and continuous diabetes-related data and created the opportunity for applying ML to automated decision support systems [8]. Combining ML-based decision support systems with the abundance of generated data has the potential to identify hypoglycemia with greater accuracy.

Conventionally, ML has been applied to detect abnormalities in blood glucose levels using physiological parameters that are highly correlated with hypoglycemia (eg, changes in brain or cardiac electrical activities) [7]. Recently, in addition to the detection of hypoglycemia, ML-based decision support systems have been proposed for predicting hypoglycemia by using various historical data (eg, series of blood glucose data, other laboratory and demographic data, verbal data in medical records, or secure messages suggesting occurrence of hypoglycemic events) [8]. Despite many reports of ML algorithms for detecting or preventing hypoglycemia, their abilities have not been comprehensively or quantitatively assessed. This meta-analysis aims to assess the current ability of ML algorithms to detect or predict hypoglycemia in patients with DM.

Methods

Protocol Registration

The study protocol has been registered in the international prospective register of systematic reviews (PROSPERO; Registration ID: CRD42020163682).

Literature Searches

We used Dialog to perform the electronic literature searches. The platform allows users to access and search 96 databases of peer-reviewed literature. Publication dates ranged from January 1, 1950, to September 14, 2020. Search terms consisted of 2 elements: (1) thesaurus and text words related to ML and (2) text words related to hypoglycemia and thesaurus terms related to glucose monitoring or blood glucose. The use of the thesaurus term was limited to 2 databases: EMBASE (EMTREE terms) and MEDLINE (MeSH terms). The above 2 elements were combined using the BOOLEAN operator “AND” (Multimedia Appendix 1). Manual searches were added to review reference lists in relevant studies. If eligible studies were obtained from the reference lists, the reference lists in those studies were also examined. Manual searches were continued until no eligible study was found in the references lists.

Study inclusion criteria were (1) all participants had DM; (2) study endpoint was hypoglycemia; (3) researchers clarified that they originally trained the ML algorithm using training data to build a model for detecting or predicting hypoglycemia or the same researchers trained the ML algorithm in a previous study; (4) the model’s performance was tested using the test data; and (5) sensitivity and specificity for detection or prediction of hypoglycemia were presented or could be calculated.

Exclusion criteria were (1) an event-based study (ie, specificity could not be estimated because nonhypoglycemia data were not included in the test data), (2) a case study (ie, training and test data were derived from only 1 patient), and (3) a 2 × 2 contingency table consisting of the number of true positives, false positives, false negatives, and false positives could not be reproduced. If studies met all of the inclusion criteria but did not allow the reproduction of a 2 × 2 contingency table, we asked the corresponding author of these studies for the total number of test data sets (N-total) and events (N-hypo) so that we could reproduce the 2 × 2 table. If the same test data were shared by 2 or more eligible studies, we chose the most updated study in which the ML algorithm was considered to show the best performance.

The outcome of meta-analyses of diagnostic or prognostic tests is the extent of consistency between an index test and a reference standard. The index test is defined as a new test that is proposed when the method for perfectly diagnosing a target condition in
all individuals does not exist or cannot be used. In this meta-analysis, it corresponded to an ML algorithm that classified the input data as either hypoglycemia or nonhypoglycemia. The reference standard is defined by a procedure that is considered the best available method for categorizing participants into having or not having a target condition. In this meta-analysis, it corresponded to methods for diagnosing hypoglycemia in clinical practice, which included measurement of glucose levels, the International Classification of Diseases (ICD) code for hypoglycemia, or experts’ subjective judgment. Evaluation item was the ability of ML algorithms to detect hypoglycemia (ie, alert to hypoglycemia coinciding with its symptoms) or the ability to predict hypoglycemia (ie, alert to hypoglycemia before its symptoms have occurred). In studies that assessed the ability for detection, data used for the index test (ie, the ML algorithm) and data used for a reference standard (ie, diagnosing hypoglycemia) had to be examined at the same time. In studies assessing predictive ability, the data input into the ML algorithm had to be examined before the diagnosis of hypoglycemia.

Data Extraction

Data were extracted by two authors (SK and KF) Disagreements were resolved by discussion with a third author (HiS). We fundamentally selected 1 datum if there were 2 or more extractable data for a set of test data in an individual study. If an individual study tested 2 or more ML classification methods or 2 or more models for 1 ML classifier, we extracted the datum related to the classifier or model that the study proposed as the best. If 2 or more different results were presented for the same model depending on the prediction window or horizon, we extracted data on the result in relation to the longest prediction window or horizon.

The following study characteristics were extracted: first author, publication year, evaluated item (ie, detecting or predicting hypoglycemia), country, type of DM (ie, type 1 or type 2), number of study participants, N-total, N-hypo, mean or range of the patients’ age, time of day of hypoglycemic events, place of supposed hypoglycemic episode (ie, experimental, in-hospital, and out-of-hospital), ML algorithm used for classification into hypoglycemia and nonhypoglycemia, threshold of glucose level for hypoglycemia, method for diagnosing hypoglycemia, method for separating the database into training and test data, and profiling data that were input into ML algorithms for performance testing.

Study Quality

To evaluate study quality, we used a revised tool to assess diagnostic accuracy of studies (QUADAS-2). The QUADAS-2 consists of 4 domains: selection of participants, index test, reference standard, and flow and timing. All 4 domains were used for assessment of risk of bias and the first 3 domains were used to assess the consensus of applicability. Each domain has 1 query in relation to the risk of bias or applicability consisting of 7 questions (Multimedia Appendix 2) [9]. A “Yes” answer was assigned 1 point.

Data Synthesis

The ability of ML algorithms to detect hypoglycemia and predict hypoglycemia was independently assessed. For data that were used to test the model’s performance, the number of true positives, false positives, true negatives, and false negatives was calculated. The set of 4 data was pooled with a hierarchical summary receiver operating characteristic (HSROC) model [10]. Indicators for the model’s performance included sensitivity, specificity, positive likelihood ratio (PLR), which is calculated as (sensitivity/[1–specificity]), and negative likelihood ratio (NLR), which is calculated as ([1–sensitivity]/specificity). Study heterogeneity was assessed by calculating I² values for PLR and NLR based on a multivariate random-effects meta-regression that considered within- and between-study correlations [11] and classifying them into quartiles (0% to <25%, low; 25% to <50%, low-to-moderate; 50% to <75%, moderate-to-high; >75%, high) [12]. Publication bias was statistically assessed as proposed by Deeks et al [13], wherein the logarithm of the diagnostic odds ratio is regressed against its corresponding inverse of the square root of the effective sample size.

Sensitivity analyses were added, and the analysis was limited to studies that shared similar characteristics in terms of the type of DM, time of day when hypoglycemia occurred, place of supposed hypoglycemic events, and the profiling data input into the ML algorithm. It is of note that at least four data sets are necessary to perform these sensitivity analyses because the HSROC model has 4 parameters: sensitivity, specificity, accuracy, and threshold. A two-sided P-value <.05 was considered statistically significant. All statistical analyses were performed using Stata 16 (StataCorp).

Results

Literature Searches

Multimedia Appendix 3 shows the flow chart of the procedure for selecting studies. Using prespecified search terms, 1226 articles were retrieved; 61 databases published at least one of the retrieved articles (Multimedia Appendix 4). Of these 1226 articles, 150 studies were selected for further review. Manual searches resulted in the addition of 32 studies for further review, making a total of 182 studies. Of these, 149 studies were subsequently excluded for various reasons. Specifically, 12 studies [14-25] presented insufficient data to allow reproduction of the 2 × 2 contingency table, although data on sensitivity and specificity were presented. We asked the authors of these studies to provide N-totals and N-hypos so that we could calculate the number of true positives, false positives, true negatives, and false negatives. However, only the author of 2 studies responded to our communication [15,25], and therefore the remaining 10 studies with insufficient data had to be excluded from the meta-analysis. Finally, 33 studies [15,20,25-55] were eligible.

Data Extraction of Study Characteristics

Table 1 shows the summary of study characteristics. Of the 33 studies, 19 studies (58%) [26-31,33,35,36,38-42,44-47,54] predicted hypoglycemia, and the remaining 14 studies (42%) detected hypoglycemia [15,20,25,32,34,37,43,48-53,55]. As much as 25 of the 33 included studies (76%) [15,20,25,27,29,30,32,35,36,38,39,41-44,46-53,55] specified type 1 as the type of DM. Type 2 DM was specified in only 3
of these studies (9%) [28,31,45] and the remaining 5 studies [33,34,37,40,54] did not specify the type of DM.

Regarding the time of day when hypoglycemic events occurred, nocturnal hypoglycemia was the most frequently reported (14 studies of the 33 included studies; 42%) [15,20,26,30,32,35,36,41,49-53]). As to the place of the supposed hypoglycemic episode, 16 of the 19 studies that predicted hypoglycemia (84%) [26-30,35,36,38-42,44-47] supposed the event took place in an out-of-hospital setting. The remaining 3 studies (16%) [31,33,54] supposed hypoglycemia occurring in an in-hospital setting. Of the 14 studies that detected hypoglycemia, 11 studies (79%) [15,20,25,32,43,48-52,55] detected hypoglycemia in an experimental setting, where hypoglycemia was induced by a hypoglycemic clamp procedure. In 20 of the 33 included studies (61%) [15,20,25,27,29,31,32,35,36,38,41,43-45,49-52,54,55], a hold-out method was used to separate the information in the database according to training and test data. Multimedia Appendix 5 shows the profiling data input into the ML algorithm for testing its performance in detecting or predicting hypoglycemia. In the majority of the 19 studies for predicting hypoglycemia (13 studies; 68%) [26-30,35,36,38,40-42,46,47], historical CGM data were input into the ML algorithm while the remaining 6 studies (32%) [31,33,39,44,45,54] did not use CGM. Of the 14 studies that detected hypoglycemia using ML, 7 studies (50%) [20,25,32,49,50,52,55] used information from electroencephalograms (EEGs) and 4 studies (29%) [15,43,51,53] used results of electrocardiography (ECG).
<table>
<thead>
<tr>
<th>Study source</th>
<th>Assessment</th>
<th>Country</th>
<th>Type of DM</th>
<th>Patients, ( n )</th>
<th>N-total(^b)</th>
<th>N-hypo(^c)</th>
<th>Mean or range of age (years)</th>
<th>Time(^d)</th>
<th>Place(^e)</th>
<th>Machine learning</th>
<th>Threshold of Hypo(^f) (mmol/L)</th>
<th>Method of Hypo detection(^g)</th>
<th>Method of separation(^h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bertachi et al [26]</td>
<td>Pre ( k )</td>
<td>Spain</td>
<td>T1D(^m)</td>
<td>10</td>
<td>124</td>
<td>39</td>
<td>32</td>
<td>Noc(^b)</td>
<td>Out(^a)</td>
<td>SVM(^y)</td>
<td>3.9</td>
<td>CGM(^i)</td>
<td>nCV(^o)</td>
</tr>
<tr>
<td>Dave et al [27]</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>112</td>
<td>637,735</td>
<td>18,233</td>
<td>13</td>
<td>N/S</td>
<td>Out(^a)</td>
<td>RF(^w)</td>
<td>3.9</td>
<td>CGM</td>
<td>HO(^p)</td>
</tr>
<tr>
<td>Elhadd et al [28]</td>
<td>Pre</td>
<td>Qatar</td>
<td>T2D(^p)</td>
<td>13</td>
<td>3918</td>
<td>172</td>
<td>51</td>
<td>N/S</td>
<td>Out</td>
<td>XG-Boost</td>
<td>Unclear</td>
<td>CGM</td>
<td>nCV</td>
</tr>
<tr>
<td>Marcus et al [29]</td>
<td>Pre</td>
<td>Israel</td>
<td>T1D</td>
<td>11</td>
<td>43,533</td>
<td>5264</td>
<td>18-39</td>
<td>N/S</td>
<td>Out</td>
<td>KRR(^x)</td>
<td>3.9</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Mosquera-Lopez et al [30], Test 1</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>10</td>
<td>117</td>
<td>17</td>
<td>34</td>
<td>Noc</td>
<td>Out</td>
<td>SVM</td>
<td>3.9</td>
<td>CGM</td>
<td>ExV</td>
</tr>
<tr>
<td>Mosquera-Lopez et al [30], Test 2</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>20</td>
<td>2706</td>
<td>258</td>
<td>35</td>
<td>Noc</td>
<td>Out</td>
<td>SVM</td>
<td>3.9</td>
<td>CGM</td>
<td>ExV</td>
</tr>
<tr>
<td>Mueller et al [31]</td>
<td>Pre</td>
<td>USA</td>
<td>T2D</td>
<td>453,487</td>
<td>90,687</td>
<td>2580</td>
<td>66</td>
<td>N/S</td>
<td>In(^a)</td>
<td>REFS</td>
<td>3.9</td>
<td>Blood/ICD</td>
<td>HO</td>
</tr>
<tr>
<td>Ngo et al [32]</td>
<td>Dec(^l)</td>
<td>Australia</td>
<td>T1D</td>
<td>8</td>
<td>135</td>
<td>53</td>
<td>12-18</td>
<td>Noc</td>
<td>Exp</td>
<td>BNN(^y)</td>
<td>3.9</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Ruan et al [33]</td>
<td>Pre</td>
<td>UK</td>
<td>N/S(^o)</td>
<td>17,658</td>
<td>3276</td>
<td>703</td>
<td>66</td>
<td>N/S</td>
<td>In</td>
<td>XG-Boost</td>
<td>3.9</td>
<td>Blood</td>
<td>nCV</td>
</tr>
<tr>
<td>Rubega et al [25]</td>
<td>Dec</td>
<td>Italy</td>
<td>T1D</td>
<td>34</td>
<td>2516</td>
<td>1258</td>
<td>55</td>
<td>N/S</td>
<td>Exp(^u)</td>
<td>NN(^z)</td>
<td>3.9</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Chen et al [34]</td>
<td>Dec</td>
<td>USA</td>
<td>N/S</td>
<td>No data</td>
<td>300</td>
<td>11</td>
<td>No data</td>
<td>N/S</td>
<td>In</td>
<td>LR(^aa)</td>
<td>N/A(^{ak})</td>
<td>Experts(^mm)</td>
<td>nCV</td>
</tr>
<tr>
<td>Guemes et al [35]</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>6</td>
<td>55</td>
<td>6</td>
<td>40-60</td>
<td>Noc</td>
<td>Out</td>
<td>SVM</td>
<td>3.9</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Jensen et al [36]</td>
<td>Pre</td>
<td>Denmark</td>
<td>T1D</td>
<td>463</td>
<td>921</td>
<td>79</td>
<td>43</td>
<td>Noc</td>
<td>Out</td>
<td>LDA(^{bb})</td>
<td>3</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Jin et al [37]</td>
<td>Dec</td>
<td>USA</td>
<td>N/S</td>
<td>No data</td>
<td>4104</td>
<td>132</td>
<td>No data</td>
<td>N/S</td>
<td>In</td>
<td>SVM</td>
<td>N/A</td>
<td>ICD(^{mn})</td>
<td>nCV</td>
</tr>
<tr>
<td>Oviedo et al [38]</td>
<td>Pre</td>
<td>Spain</td>
<td>T1D</td>
<td>10</td>
<td>1447</td>
<td>420</td>
<td>41</td>
<td>Pos(^i)</td>
<td>Out</td>
<td>SVM</td>
<td>3.9</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Reddy et al [39]</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>55</td>
<td>90</td>
<td>29</td>
<td>33</td>
<td>Ex</td>
<td>Out</td>
<td>RF</td>
<td>3.9</td>
<td>Blood</td>
<td>ExV</td>
</tr>
<tr>
<td>Seo et al [40]</td>
<td>Pre</td>
<td>Korea</td>
<td>N/S</td>
<td>104</td>
<td>7052</td>
<td>412</td>
<td>52</td>
<td>Pos</td>
<td>Out</td>
<td>RF</td>
<td>3.9</td>
<td>CGM</td>
<td>nCV</td>
</tr>
<tr>
<td>Study source</td>
<td>Assessment</td>
<td>Country</td>
<td>Type of DM</td>
<td>Patients, n</td>
<td>N-total&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N-hypo&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Mean or range of age (years)</td>
<td>Time&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Place&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Machine learning</td>
<td>Threshold of Hypo&lt;sup&gt;f&lt;/sup&gt; (mmol/L)</td>
<td>Method of Hypo detection&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Method of separation&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Arthur et al [41]</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>6</td>
<td>51</td>
<td>6</td>
<td>40-60</td>
<td>Noc</td>
<td>Out</td>
<td>ANN&lt;sup&gt;cc&lt;/sup&gt;</td>
<td>3.9</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Tof-fanin et al [42]</td>
<td>Pre</td>
<td>Italy</td>
<td>T1D</td>
<td>20</td>
<td>7096</td>
<td>36</td>
<td>46</td>
<td>N/S</td>
<td>Out</td>
<td>I-MPC&lt;sup&gt;dd&lt;/sup&gt;</td>
<td>3.9</td>
<td>CGM</td>
<td>ExV</td>
</tr>
<tr>
<td>Ling et al [43]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>16</td>
<td>269</td>
<td>55</td>
<td>15</td>
<td>N/S</td>
<td>Exp</td>
<td>FNN&lt;sup&gt;ee&lt;/sup&gt;</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Sampath et al [44], DIA&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Pre</td>
<td>Ukraine</td>
<td>T1D</td>
<td>34</td>
<td>150</td>
<td>40</td>
<td>18-65</td>
<td>Noc</td>
<td>Out</td>
<td>RA&lt;sup&gt;ff&lt;/sup&gt;</td>
<td>3.9</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Sampath et al [44], Child&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Pre</td>
<td>Ukraine</td>
<td>T1D</td>
<td>179</td>
<td>476</td>
<td>222</td>
<td>3-16</td>
<td>Noc</td>
<td>Out</td>
<td>RA&lt;sup&gt;ff&lt;/sup&gt;</td>
<td>3.9</td>
<td>Blood</td>
<td>ExV</td>
</tr>
<tr>
<td>Sudharsan et al [45]</td>
<td>Pre</td>
<td>USA</td>
<td>T2D</td>
<td>Unclear</td>
<td>839</td>
<td>428</td>
<td>No data</td>
<td>N/S</td>
<td>Out</td>
<td>RF&lt;sup&gt;gg&lt;/sup&gt;</td>
<td>3.9</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Eljil [46]</td>
<td>Pre</td>
<td>UAE</td>
<td>T1D</td>
<td>10</td>
<td>667</td>
<td>100</td>
<td>25</td>
<td>N/S</td>
<td>Out</td>
<td>BAG&lt;sup&gt;gg&lt;/sup&gt;</td>
<td>3.3</td>
<td>CGM</td>
<td>nCV</td>
</tr>
<tr>
<td>Plis et al [47]</td>
<td>Pre</td>
<td>USA</td>
<td>T1D</td>
<td>2</td>
<td>5816</td>
<td>152</td>
<td>No data</td>
<td>N/S</td>
<td>Out</td>
<td>SVM&lt;sup&gt;hh&lt;/sup&gt;</td>
<td>3.9</td>
<td>CGM</td>
<td>ExV</td>
</tr>
<tr>
<td>Jensen et al [48]</td>
<td>Dec</td>
<td>Denmark</td>
<td>T1D</td>
<td>10</td>
<td>1267</td>
<td>160</td>
<td>44</td>
<td>N/S</td>
<td>Exp</td>
<td>SEP-COR&lt;sup&gt;iih&lt;/sup&gt;</td>
<td>3.9</td>
<td>Blood</td>
<td>LOO</td>
</tr>
<tr>
<td>Jensen et al [48]</td>
<td>Dec</td>
<td>Denmark</td>
<td>T1D</td>
<td>10</td>
<td>1267</td>
<td>160</td>
<td>44</td>
<td>N/S</td>
<td>Exp</td>
<td>+ SVM&lt;sup&gt;iih&lt;/sup&gt;</td>
<td>3.9</td>
<td>Blood</td>
<td>LOO</td>
</tr>
<tr>
<td>Nguyen et al [49]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>5</td>
<td>144</td>
<td>76</td>
<td>12-18</td>
<td>Noc</td>
<td>Exp</td>
<td>FNN</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Nguyen et al [50]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>5</td>
<td>44</td>
<td>20</td>
<td>12-18</td>
<td>Noc</td>
<td>Exp</td>
<td>ANN</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Nuryani et al [51]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>5</td>
<td>575</td>
<td>133</td>
<td>16</td>
<td>Noc</td>
<td>Exp</td>
<td>PSO&lt;sup&gt;jj&lt;/sup&gt; + SVM</td>
<td>Unclear</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Chan et al [15]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>16</td>
<td>100</td>
<td>52</td>
<td>15</td>
<td>Noc</td>
<td>Exp</td>
<td>FNN</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Ling et al [52]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>5</td>
<td>27</td>
<td>8</td>
<td>16</td>
<td>Noc</td>
<td>Exp</td>
<td>Fuzzy SVM</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
<tr>
<td>Nguyen and Jones [20]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>6</td>
<td>79</td>
<td>27</td>
<td>12-18</td>
<td>Noc</td>
<td>Exp</td>
<td>BNN</td>
<td>3.3</td>
<td>Blood</td>
<td>HO</td>
</tr>
<tr>
<td>Skladnev et al [53]</td>
<td>Dec</td>
<td>Australia</td>
<td>T1D</td>
<td>32</td>
<td>52</td>
<td>11</td>
<td>16</td>
<td>Noc</td>
<td>In</td>
<td>FNN</td>
<td>3.9</td>
<td>Blood</td>
<td>ExV</td>
</tr>
<tr>
<td>Zhang et al [54]</td>
<td>Pre</td>
<td>USA</td>
<td>N/S</td>
<td>1004</td>
<td>1114</td>
<td>556</td>
<td>No data</td>
<td>N/S</td>
<td>In</td>
<td>DT&lt;sup&gt;jj&lt;/sup&gt;</td>
<td>3.3</td>
<td>CGM</td>
<td>HO</td>
</tr>
</tbody>
</table>

<sup>a</sup> Assessment method used by each study.  
<sup>b</sup> Number of patients included.  
<sup>c</sup> Machine learning algorithm for classification of hypo.  
<sup>d</sup> Time of measurement.  
<sup>e</sup> Place of measurement.  
<sup>f</sup> Threshold of hypo.  
<sup>g</sup> Method of hypo detection.  
<sup>h</sup> Method of separation.  
<sup>1</sup> DIA: Diabetes Information Association.  
<sup>2</sup> Blood: Blood glucose measurement.  
<sup>3</sup> Exp: Experimental dataset.  
<sup>4</sup> In: Invasive method.  
<sup>5</sup> Out: Non-invasive method.  
<sup>6</sup> SVM: Support vector machine.  
<sup>7</sup> ANN: Artificial neural network.  
<sup>8</sup> FNN: Feedforward neural network.  
<sup>9</sup> RA: Region of interest analysis.  
<sup>10</sup> LOO: Leave-one-out cross-validation.  
<sup>11</sup> BNN: Bayesian neural network.  
<sup>12</sup> Fuzzy SVM: Fuzzy support vector machine.  
<sup>13</sup> DT: Decision tree.  
<sup>14</sup> PSO: Particle swarm optimization.  
<sup>15</sup> SVM: Support vector machine.  
<sup>16</sup> SVM: Support vector machine.  
<sup>17</sup> SVM: Support vector machine.  
<sup>18</sup> SVM: Support vector machine.  
<sup>19</sup> SVM: Support vector machine.  
<sup>20</sup> SVM: Support vector machine.  
<sup>21</sup> SVM: Support vector machine.  
<sup>22</sup> SVM: Support vector machine.  
<sup>23</sup> SVM: Support vector machine.  
<sup>24</sup> SVM: Support vector machine.  
<sup>25</sup> SVM: Support vector machine.  
<sup>26</sup> SVM: Support vector machine.  
<sup>27</sup> SVM: Support vector machine.  
<sup>28</sup> SVM: Support vector machine.  
<sup>29</sup> SVM: Support vector machine.  
<sup>30</sup> SVM: Support vector machine.  
<sup>31</sup> SVM: Support vector machine.  
<sup>32</sup> SVM: Support vector machine.  
<sup>33</sup> SVM: Support vector machine.  
<sup>34</sup> SVM: Support vector machine.  
<sup>35</sup> SVM: Support vector machine.  
<sup>36</sup> SVM: Support vector machine.  
<sup>37</sup> SVM: Support vector machine.  
<sup>38</sup> SVM: Support vector machine.  
<sup>39</sup> SVM: Support vector machine.  
<sup>40</sup> SVM: Support vector machine.  
<sup>41</sup> SVM: Support vector machine.  
<sup>42</sup> SVM: Support vector machine.  
<sup>43</sup> SVM: Support vector machine.  
<sup>44</sup> SVM: Support vector machine.  
<sup>45</sup> SVM: Support vector machine.  
<sup>46</sup> SVM: Support vector machine.  
<sup>47</sup> SVM: Support vector machine.  
<sup>48</sup> SVM: Support vector machine.  
<sup>49</sup> SVM: Support vector machine.  
<sup>50</sup> SVM: Support vector machine.  
<sup>51</sup> SVM: Support vector machine.  
<sup>52</sup> SVM: Support vector machine.  
<sup>53</sup> SVM: Support vector machine.  
<sup>54</sup> SVM: Support vector machine.  

http://diabetes.jmir.org/2021/1/e22458/
<table>
<thead>
<tr>
<th>Study Source</th>
<th>Assessment&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Country</th>
<th>Type of DM</th>
<th>Patients, N</th>
<th>N-total&lt;sup&gt;b&lt;/sup&gt;</th>
<th>N-hypo&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Mean or range of age (years)</th>
<th>Time&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Place&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Machine learning</th>
<th>Threshold of Hypo&lt;sup&gt;f&lt;/sup&gt; (mmol/L)</th>
<th>Method of Hypo detection&lt;sup&gt;g&lt;/sup&gt;</th>
<th>Method of separation&lt;sup&gt;h&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iaione and Marques [55]</td>
<td>Dec</td>
<td>Brazil</td>
<td>T1D</td>
<td>8</td>
<td>1990</td>
<td>995</td>
<td>35</td>
<td>Mor&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Exp</td>
<td>ANN</td>
<td>3.3</td>
<td>Blood</td>
<td>HO</td>
</tr>
</tbody>
</table>

<sup>a</sup> Ability for which the machine learning algorithm was assessed.
<sup>b</sup> N-total: total number of data included in test data.
<sup>c</sup> N-hypo: total number of hypoglycemic episodes included in the test data.
<sup>d</sup> Time of day when hypoglycemia occurred.
<sup>e</sup> Place of supposed hypoglycemic episode.
<sup>f</sup> Threshold of glucose level that was used to diagnose hypoglycemia.
<sup>g</sup> Method for separating training and test data.
<sup>h</sup> Method used for diagnosing hypoglycemia.

<sup>i</sup> DIA: DIAdvisor.
<sup>j</sup> Child: ChildrenData.
<sup>k</sup> Pre: predicting hypoglycemia.
<sup>l</sup> Dec: detecting hypoglycemia.
<sup>m</sup> T1D: type 1 diabetes mellitus.
<sup>n</sup> T2D: type 2 diabetes mellitus.
<sup:o</sup> N/S: not specified.
<sup>p</sup> NOC: nocturnal hypoglycemia.
<sup>q</sup> Pos: postprandial.
<sup>r</sup> Mor: hypoglycemia during morning.
<sup>s</sup> Out: out-of-hospital setting.
<sup>t</sup> In: in-hospital setting.
<sup>u</sup> Exp: experimental setting (ie, hypoglycemia is induced by injection of insulin. Exercise or drug intervention is included in out of hospital setting).
<sup>v</sup> SVM: support vector machine.
<sup>w</sup> RF: random forest.
<sup>x</sup> KRR: Kernel Ridge Regression.
<sup>y</sup> BNN: Bayesian neural network.
<sup>z</sup> NN: neural network.
<sup>aa</sup> LR: logistic regression.
<sup>bb</sup> LDA: linear discriminant analysis.
<sup>cc</sup> ANN: artificial neural network.
<sup>dd</sup> MPC: individual model-based predictive control.
<sup>ee</sup> FNN: fuzzy neural network.
<sup>ff</sup> RA: ranking aggregation algorithms.
<sup>gg</sup> BAG: bagging (bootstrap aggregating).
<sup>hh</sup> SEPCOR: separability and correlation analysis.
<sup>ii</sup> PSO: particle swarm optimization.
<sup jj</sup> DT: decision tree.
<sup>kk</sup> N/A: Not applicable.
<sup>ll</sup> CGM: continuous glucose monitoring.
<sup>mm</sup> Experts' subjective judgment.
<sup>nn</sup> ICD: International Classification of Diseases.
<sup>oo</sup> nCV: n-fold cross-validation.
<sup>pp</sup> HO: hold-out method.
<sup>qq</sup> LOO: leave-one-out cross-validation.
Assessment of Study Quality

Multimedia Appendix 6 shows the results of study quality assessments using QUADAS-2. Mean score (SD) was 5.6 (1.1), which corresponded to 80% of full marks (=7). The applicability of the reference test was evaluated to be low in 61% of the 33 included studies (20 studies) because hypoglycemia was not diagnosed by measuring blood glucose levels or ICD codes but by CGM (ie, glucose levels in blood are indirectly estimated from those in interstitial tissue) (19 studies) [15,26-30,35,38,40-43,46,47,49-52,54] or experts’ subjective judgement (1 study) [34]. The 2 factors were mainly responsible for lowering the study quality. We considered that the threshold of hypoglycemia in the index test was not specified in 7 studies, which used the cross-validation method [26,28,33,34,37,40,46], and 1 study, which used the leave-one-out method to separate test data from training data [48].

Data Synthesis

Ability for Detection of Hypoglycemia Using ML Algorithms

Figure 1 shows the HSROC curve and pooled estimates of sensitivity and specificity based on the 14 studies that assessed the ability of the ML algorithm to detect hypoglycemia. The pooled estimates (95% CI) were 0.79 (0.75-0.83) for sensitivity and 0.80 (0.64-0.91) for specificity. The pooled estimates (95% CI) of PLR and NLR were 2.20 (1.46-3.32) and 0.37 (0.28-0.49), respectively. Between-study heterogeneity expressed as $I^2$ was high both for PLR (98%; 95% CI 95%-99%) and NLR (80%; 95% CI 50%-90%). Statistically significant publication bias was detected ($P=.15$).

Figure 1. Hierarchical summary receiver-operating characteristic (HSROC) curve for detection of hypoglycemia using machine learning algorithms. Circles indicate study-specific sensitivity and specificity for each of the 14 included studies. The size of each circle is proportional to study sample size. The pooled point estimates of sensitivity and specificity are plotted in a filled square.

We conducted several sensitivity analyses using a portion of the above 14 studies that had 1 study characteristic in common. It was not apparent that any of the sensitivity analyses showed results different from the overall analysis. Limiting the analyses to 12 studies [15,20,25,32,43,48-53,55] that specified type 1 as the DM type, pooled sensitivity, specificity, PLR, and NLR were 0.78 (95% CI 0.73-0.82), 0.71 (95% CI 0.60-0.79), 2.65 (95% CI 1.88-3.72), and 0.26 (95% CI 0.19-0.36), respectively. When analyses were limited to the 7 studies that detected nocturnal hypoglycemia using ML algorithms [15,20,49-53], the pooled estimates (95% CI) were 0.75 (0.70-0.80) for sensitivity, 0.65 (0.55-0.74) for specificity, 2.14 (1.67-2.76) for PLR, and 0.38 (0.30-0.48) for NLR. With analyses of the 11 studies that detected hypoglycemia in an experimental setting, pooled sensitivity, specificity, PLR, and NLR were 0.78 (95% CI 0.73-0.82), 0.71 (95% CI 0.60-0.80), 2.66 (95% CI 1.84-3.85), and 0.31 (0.24-0.41), respectively. The pooled estimate (95% CI) was 0.78 (0.71-0.84) for sensitivity, 0.67 (0.55-0.77) for specificity, 2.39 (1.63-3.50) for PLR, and 0.33 (0.22-0.48) for NLR when the analysis was limited to 7 studies that used EEG abnormalities for detecting hypoglycemia. These estimations were similar when limited to 4 studies that used ECG abnormalities for detection of hypoglycemia: pooled estimate (95% CI) was 0.76 (0.67-0.82) for sensitivity; 0.67 (0.54-0.78) for specificity; 2.31 (1.65-3.23) for PLR; and 0.36 (0.28-0.47) for NLR.

Ability to Predict Hypoglycemia Using ML Algorithms

Figure 2 shows the HSROC curve for predicting hypoglycemia based on the 19 studies that assessed the predictive ability for
hypoglycemia. The point estimates (95% CI) were 0.80 (0.72-0.86) for sensitivity, 0.92 (0.87-0.96) for specificity, 10.42 (5.82-18.65) for PLR, and 0.22 (0.15-0.31) for NLR. Extremely high between-study heterogeneity was observed for both PLR ($I^2$ [95% CI] 100% [100%-100%]) and NLR ($I^2$ [95% CI] 99% [98%-100%]). Publication bias was not statistically significant ($P=0.68$).

**Figure 2.** Hierarchical summary receiver-operating characteristic (HSROC) curve for prediction of hypoglycemia using machine learning algorithms. Circles indicate study-specific sensitivity and specificity for each of the 19 included studies. The size of each circle is proportional to study sample size. The pooled point estimates of sensitivity and specificity are plotted in a filled square.

When the analyses were limited to 13 studies that specified type 1 as the DM type [26,27,29,30,35,36,38,39,41,42,44,46,47], the pooled estimates (95% CI) were 0.77 (0.67-0.85) for sensitivity, 0.92 (0.84-0.96) for specificity, 9.82 (4.58-21.04) for PLR, and 0.25 (0.16-0.38) for NLR. In the analyses of 7 studies that specified night as the time of hypoglycemic events [26,30,31,35,36,41,44], the predictive ability was low compared with that of the overall analysis—pooled estimate (95% CI): 0.74 (0.65-0.82) for sensitivity, 0.81 (0.72-0.88) for specificity, 3.98 (2.64-6.00) for PLR, and 0.31 (0.23-0.43) for NLR. Relatively high sensitivity and low NLR were observed in the 13 studies that used CGM historical data for predicting hypoglycemia—pooled estimate (95% CI): 0.82 (0.71-0.90) for sensitivity, 0.92 (0.83-0.97) for specificity, 10.41 (4.52-24.01) for PLR, and 0.19 (0.12-0.32) for NLR—compared with 6 studies that did not use CGM—pooled estimate (95% CI): 0.76 (0.66-0.84) for sensitivity, 0.92 (0.88-0.95) for specificity, 10.14 (6.13-16.77) for PLR, and 0.26 (0.17-0.38) for NLR. After excluding 3 studies [31,33,54] that showed that the supposed hypoglycemic events occurred in-hospital, the pooled estimates (95% CI) of the 16 studies with such events occurring in an out-of-hospital setting were 0.82 (0.74-0.88) for sensitivity, 0.92 (0.85-0.96) for specificity, 10.58 (5.44-20.55) for PLR, and 0.20 (0.13-0.39) for NLR.

**Discussion**

**Principal Findings**

Overall, the PLR and NLR of ML algorithms for detecting hypoglycemia were 4.05 and 0.26, respectively. These estimates were almost unchanged throughout several sensitivity analyses that were limited to studies that shared 1 characteristic in common. According to the Users’ Guide to Medical Literature with regard to diagnostic tests [56], the PLR should be 5 or more to moderately increase the probability of persons having or developing a disease and the NLR should be 0.2 or less to moderately decrease the probability of having or developing a disease after taking the index test. In summary, the current ML algorithms had insufficient ability to detect the occurrence of hypoglycemia. However, that would not mean that ECG or EEG monitoring in combination with ML, which was the case with 79% (11/14) of the included studies, was useless in detecting hypoglycemia. For example, for patients with both DM and high cardiovascular risk, in particular, those who are vulnerable to cardiac arrhythmias, using ECGs for detecting hypoglycemia is useful considering that a hypoglycemia-induced arrhythmia could contribute to increased cardiovascular mortality [57]. Similarly, for patients with repeated episodes of hypoglycemia, the combination of ML and EEG was indicated to be beneficial to prevent hypoglycemia-induced neuroglycopenia resulting in cognitive impairment and ultimately death, because blood glucose levels alone do not appear to predict that condition [58].
Thus, the clinical applicability of these devices should be evaluated by the individual’s risk of hypoglycemia and its related arrhythmia and neuroglycopenia as well as the overall ability of algorithms for ML.

The overall sensitivity, specificity, PLR, and NLR for predicting hypoglycemia were 0.80, 0.92, 10.42, and 0.22, respectively. Applying the above described guidelines for diagnostic tests to these results, it is worth considering the use of current ML algorithms as a tool for alerting patients to impending hypoglycemic events. In addition, it is considered that a test with a PLR over 10 has a particularly strong power to alter posttest probability of the targeted disease compared with pretest probability [56]. If a positive test result were to be received, patients with DM who are administered hypoglycemic treatments would be strongly recommended to pay more attention to the possibility of impending hypoglycemic events than they would before receiving the predictive test for hypoglycemia. However, considering that the PLR and NLR values indicate relative risk (ie, risk of disease at posttest compared with that at pretest), the accuracy of predictive ability depends on patients’ risk of hypoglycemia in daily life. For example, even a less than 10% false-positive rate (8% in this meta-analysis) may be acceptable in patients at high risk of hypoglycemia but not in low-risk individuals due to too frequent false alarms. In such a case, there is fear that these patients will ignore the alarms and therefore miss the opportunity to take corrective action when the alarm is indeed true [59]. It is emphasized that the utility of ML algorithm depends on the extent of the patient’s risk of hypoglycemia. In addition, as indicated in the “Results” section, there was high between-study heterogeneity among studies. Specifically, when limiting analyses to the studies that predicted nocturnal hypoglycemia, the predictive ability was insufficient (pooled estimate: 3.98 for PLR; 0.31 for NLR). Considering that nocturnal hypoglycemia is the most common type of hypoglycemia among all hypoglycemic episodes [60], continued research is needed for further development of ML algorithms to predict hypoglycemia.

Several limitations of this meta-analysis should be addressed. First, the principal major limitation is the pooling of studies among which there was much variability in the type of DM, profiling data for detecting or predicting hypoglycemia, time of day when hypoglycemic events occurred, setting of supposed hypoglycemic events, and ML classification methods. In particular, although the ability for predicting hypoglycemia depended largely on the ML classification methods [33], this meta-analysis did not consider the difference in the test performance among various ML methods. Instead, the meta-analysis focused on ML’s comprehensive ability across studies using data in relation to the best model in each study, if 2 or more models existed, rather than comparisons among 2 or more models within 1 study. Given that generalization of evidence is among the most important roles in all meta-analyses, the issue of the variation in ML methods, in particular, the difference between old and new ML techniques, might be beyond the scope of this meta-analysis. Nevertheless, it should be emphasized that successful application of ML lies in the correct understanding of the advantages and disadvantages of different ML methods. Second, only 3 studies exclusively targeted patients with type 2 DM. With the increasing use of insulin to treat type 2 DM in the elderly, the prevalence of hypoglycemia is likely to escalate. In addition, the response to hypoglycemia is different between type 1 and type 2 DM [61]. Future studies should aim to develop and validate ML algorithms for detecting or predicting hypoglycemia in type 2 DM. Third, in most of the included studies, the ML classification models were developed in an experimental setting or by using previously recorded data as training and testing data instead of live data. Future studies need to train and test the algorithm on data from DM patients in everyday clinical practice to determine feasibility.

Conclusion
Overall, current ML algorithms have insufficient ability to detect ongoing hypoglycemia and considerable ability to predict hypoglycemia in patients with DM receiving hypoglycemic treatments. However, the clinical applicability of these ML algorithms should be evaluated according to patients’ risk profiles such as for hypoglycemia and its associated complications (eg, arrhythmia, neuroglycopenia) as well as the average ability of the ML algorithm. Continued research is required to further develop ML algorithms to enhance their feasibility, considering the inaccuracy of CGM in the hypoglycemic range, the increased prevalence of hypoglycemia in the elderly, and increasing evidence for the effectiveness of tight glycemic control in preventing microvascular complications [62].

Acknowledgments
All authors thank Ms Haga and Ms Chino in Niigata University for their excellent secretarial work. SK was financially supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (ID: 19K12840). The sponsor had no influence over the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy in this meta-analysis.
Multimedia Appendix 2
Study quality assessment using the quality assessment of diagnostic accuracy studies (QUADUS-2).

Multimedia Appendix 3
Study flow in this meta-analysis.

Multimedia Appendix 4
Databases which published articles that were retrieved by the search terms (see Appendix 1).

Multimedia Appendix 5
Profiling data input into ML algorithm for testing its performance.

Multimedia Appendix 6
Results of assessing study quality using revised tool for the quality assessment of diagnostic accuracy studies (QUADUS-2). The criterion corresponding to each domain (D) and signaling question (SQ) is indicated in Appendix 2.

References


43. Ling SH, San PP, Nguyen HT. Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes. ISA Trans 2016 Sep;64:440-446. [doi: 10.1016/j.isatra.2016.05.008] [Medline: 27311357]


51. Ling S, Nuryani N, Nguyen H. editors. 2010 Jul 18 Presented at: Hypoglycemia detection for type 1 diabetic patients based on ECG parameters using Fuzzy Support Vector Machine. the International Joint Conference on Neural Networks (IJCNN); Barcelona, Spain; 2010; Barcelona, Spain. [doi: 10.1109/ijcnn.2010.5596916]


http://diabetes.jmir.org/2021/1/e22458/

Abbreviations
CGM: continuous glucose monitoring
DM: diabetes mellitus
HSROC: hierarchical summary receiver operating characteristic
ICD: International Classification of Diseases
ML: machine learning
N-hypo: total number of events
NLR: negative likelihood ratio
N-total: total number of test data sets
PLR: positive likelihood ratio

©Satoru Kodama, Kazuya Fujihara, Haruka Shiozaki, Chika Horikawa, Mayuko Harada Yamada, Takaaki Sato, Yuta Yaguchi, Masahiko Yamamoto, Masaru Kitazawa, Midori Iwanaga, Yasuhiro Matsubayashi, Hirohito Sone. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 29.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Experiences of Young People and Their Caregivers of Using Technology to Manage Type 1 Diabetes Mellitus: Systematic Literature Review and Narrative Synthesis

Nicola Brew-Sam1*, BA, MA, PhD; Madhur Chhabra1*, BDS, MPH; Anne Parkinson1, BA (Hons), PhD, AFHEA; Kristal Hannan1; Ellen Brown1; Lachlan Pedley1; Karen Brown1,2, BA, RN; Kristine Wright1,2, BSc, RN, CDE; Elizabeth Pedley1,2, RN, RM; Christopher J Nolan2,3,4, BMedSci, MBBS, PhD, FRACP; Christine Phillips1, BMedSc, MBBS, MA, MPH, FRACGP, MD; Hanna Suominen5,6, MSc, PhD, MEDL; Antonio Tricoli4,8, BSc, MSc, PhD; Jane Desborough1, RN, RM, MPH, PhD

1Department of Health Services Research and Policy, Research School of Population Health, College of Health and Medicine, Australian National University, Canberra, Australia
2Canberra Health Services, Canberra, Australia
3ANU Medical School, College of Health and Medicine, Australian National University, Canberra, Australia
4The John Curtin School of Medical Research, College of Health and Medicine, Australian National University, Canberra, Australia
5School of Computing, College of Engineering and Computer Science, Australian National University, Canberra, Australia
6Department of Computing, University of Turku, Turku, Finland
7Data61, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
8Nanotechnology Research Lab, Research School of Chemistry, College of Science, Australian National University, Canberra, Australia
* these authors contributed equally

Corresponding Author:
Nicola Brew-Sam, BA, MA, PhD
Department of Health Services Research and Policy
Research School of Population Health, College of Health and Medicine
Australian National University
Building 62 Mills Rd, Acton ACT
Canberra, 2601
Australia
Phone: 61 0480238211
Email: nbrewsam@gmail.com

Abstract

Background: In the last decade, diabetes management has begun to transition to technology-based care, with young people being the focus of many technological advances. Yet, detailed insights into the experiences of young people and their caregivers of using technology to manage type 1 diabetes mellitus are lacking.

Objective: The objective of our study was to describe the breadth of experiences and perspectives on diabetes technology use among children and adolescents with type 1 diabetes mellitus and their caregivers.

Methods: This systematic literature review used integrated thematic analysis to guide a narrative synthesis of the included studies. We analyzed the perspectives and experiences of young people with type 1 diabetes mellitus and their caregivers reported in qualitative studies, quantitative descriptive studies, and studies with a mixed methods design.

Results: Seventeen articles met the inclusion criteria, and they included studies on insulin pump, glucose sensors, and remote monitoring systems. The following eight themes were derived from the analysis: (1) expectations of the technology prior to use, (2) perceived impact on sleep and overnight experiences, (3) experiences with alarms, (4) impact on independence and relationships, (5) perceived usage impact on blood glucose control, (6) device design and features, (7) financial cost, and (8) user satisfaction. While many advantages of using diabetes technology were reported, several challenges for its use were also reported, such as cost, the size and visibility of devices, and the intrusiveness of alarms, which drew attention to the fact that the user had type 1 diabetes mellitus. Continued use of diabetes technology was underpinned by its benefits outweighing its challenges, especially among younger people.
Conclusions: Diabetes technologies have improved the quality of life of many young people with type 1 diabetes mellitus and their caregivers. Future design needs to consider the impact of these technologies on relationships between young people and their caregivers, and the impact of device features and characteristics such as size, ease of use, and cost.

(JMIR Diabetes 2021;6(1):e20973) doi:10.2196/20973

KEYWORDS

Type 1 diabetes mellitus; diabetes; children; adolescents; technology; self-management; experiences; perspectives; systematic review

Introduction

Background

Type 1 diabetes mellitus (T1DM) is a chronic autoimmune disease that results in elevated blood glucose levels due to destruction of insulin-producing pancreatic islet β cells [1]. It is frequently diagnosed among children and adolescents, with the peak age group of diagnosis being 10 to 19 years [2,3]. Globally, the prevalence of T1DM among children and adolescents equates to over 1 million people currently affected [4]. Continuous glucose monitoring (CGM) has been found to have a positive impact on young people’s health-related quality of life [5,6]; therefore, technology-supported care approaches specifically for children and adolescents continue to be developed and improved [7]. Further adaptation of diabetes technology for use by young people and their caregivers can optimize diabetes management and outcomes from an early age. Insight into the experiences of young people and their caregivers of using devices to manage T1DM is essential to guide device developers and health care professionals to optimize the use and function of these technologies [8,9].

Diabetes Management in Youth

Disease management at an early age requires interdisciplinary care coordination between the child, the parents/family, the health care professional team [10], and others involved in care, such as teachers [11]. The diagnosis of diabetes at a young age is frequently accompanied by psychological stress in both the child or adolescent and parents related to the disease management demands (24 hours a day, 7 days a week), including the integration of complex treatment regimens [12] and fear of the consequences of poor blood glucose control, particularly hypoglycemia [13,14]. For adolescents, diabetes management can be a major challenge as a consequence of growing independence from parents, increasing complexity of daily activities (eg, managing diabetes technology), the added psychological demands associated with this age including peer pressure [11], and the pubertal physiological changes in the body.

Technology for Diabetes Management

To achieve optimal blood glucose control, adolescents with T1DM have to manage the following three key components: (1) glucose monitoring, (2) insulin delivery, and (3) means of communication between (1) and (2). Exogenous insulin administration into subcutaneous tissues by insulin injection or infusion by pump is informed by measurement of either blood glucose or subcutaneous interstitial fluid glucose. Such treatment is necessary to avoid short-term complications (eg, hypoglycemic events and diabetic ketoacidosis) and long-term complications (eg, diabetic retinopathy and nephropathy) [1,15]. For glucose monitoring, the choices include finger stick blood sampling for self-monitoring of blood glucose (SMBG) and/or continuous subcutaneous interstitial fluid glucose measurement with real-time access using CGM systems and/or intermittent access using flash glucose monitoring (FGM) systems. The choices for insulin delivery are multiple dose injections or continuous subcutaneous insulin infusion (CSI) by pump [16]. All combinations of glucose monitoring and insulin delivery devices are used in current practice [17]. Until recently, there were no direct electronic means of communication between the glucose monitoring and insulin delivery systems, such that a young person with diabetes or a parent/caregiver would need to make all decisions. New technology, however, has brought new means of communication between glucose sensing devices, people with diabetes, and insulin delivery systems [16]. Safety features, such as “suspend before low,” and glucose sensing-insulin infusion closed loop (CL) systems, can now be used. Hybrid closed loop (HCL) systems, in which the operating person provides some information into the otherwise CL system, such as carbohydrate intake amount that triggers an insulin bolus, are now commercially available. Table 1 provides a comprehensive technology overview [18-25].

Previous reviews on diabetes technology have mostly focused on the effectiveness or efficacy of the technology in adult populations [26-28], with some also including youth [29]. While various studies have focused on experiences with diabetes technology and particularly experiences with technology in young people with T1DM, reviews of such study findings are still lacking. Therefore, this systematic integrative review aimed to describe the breadth of experiences and perspectives on diabetes technology use among adolescents with T1DM and their caregivers.
The search string included the following keywords: (“Type 1 diabetes” OR “insulin dependent diabetes mellitus” OR “juvenile diabetes”) AND (“self manage*” OR “self measure*” OR “self monitor*”) AND (adolescent OR children) AND experience*. We did not use the term “technology” or a similar term in the search string because this limited the results considerably (a comparison was conducted). The reference lists of included studies were searched to include studies that did not appear in the database search. The Cochrane software Covidence [32] was used to assist in the systematic review process from screening to data extraction.

### Inclusion/Exclusion Criteria

Owing to the lack of age specification in many studies, we included studies with participants aged 12 to 25 years to ensure we captured adolescents, who were our primary interest. Studies that focused on parents’ or caregivers’ experiences of caring for a young person with T1DM were also included. We included peer-reviewed studies conducted in any country and in English language from 2009 to early 2019. We excluded randomized controlled trials (RCTs) owing to the integrative narrative scope of the review, which aimed to understand experiences rather than efficacy and effectiveness of technology. Other systematic reviews, conference abstracts, and grey literature were excluded.

### Screening and Quality Assessment

Selected studies were reviewed independently by two researchers, based first on the title and abstract and then on full-text review. Conflicts were resolved through discussion with a third independent reviewer. A full-text quality appraisal

### Review Design

This systematic literature review was based on the design synthesis methods of the Evidence for Policy and Practice Information Centre (EPPI-Centre) [30] and the integrative review methodology described by Whittemore and Knaf [31]. Integrative reviews enable the synthesis of data from diverse sources (qualitative and quantitative) to provide a broad and holistic understanding of the subjective and objective elements of a topic, including context, processes, and outcomes [31]. Integrated thematic analysis of data guided a narrative synthesis of the results. Data from qualitative, quantitative, and mixed methods studies were included in this narrative synthesis. The review was registered with PROSPERO (registration number: CRD42019125351).

### Patient and Public Involvement

In the true spirit of patient and public involvement in research, our team included academics, clinicians, three young people with T1DM, and two of their parents. All team members have contributed to this review, including identifying appropriate search terms, assisting with data extraction and data analysis, and providing comments on various drafts of the manuscript.

### Search Strategy

We searched PubMed, CINAHL, MEDLINE, Scopus, ProQuest, and Web of Science (search in title/abstract). The search string included the following keywords: (“Type 1 diabetes” OR “insulin dependent diabetes mellitus” OR “juvenile diabetes”) AND (“self manage*” OR “self measure*” OR “self monitor*”) AND (adolescent OR children) AND experience*. We did not use the term “technology” or a similar term in the search string because this limited the results considerably (a comparison was conducted). The reference lists of included studies were searched to include studies that did not appear in the database search. The Cochrane software Covidence [32] was used to assist in the systematic review process from screening to data extraction.

### Inclusion/Exclusion Criteria

Owing to the lack of age specification in many studies, we included studies with participants aged 12 to 25 years to ensure we captured adolescents, who were our primary interest. Studies that focused on parents’ or caregivers’ experiences of caring for a young person with T1DM were also included. We included peer-reviewed studies conducted in any country and in English language from 2009 to early 2019. We excluded randomized controlled trials (RCTs) owing to the integrative narrative scope of the review, which aimed to understand experiences rather than efficacy and effectiveness of technology. Other systematic reviews, conference abstracts, and grey literature were excluded.

### Screening and Quality Assessment

Selected studies were reviewed independently by two researchers, based first on the title and abstract and then on full-text review. Conflicts were resolved through discussion with a third independent reviewer. A full-text quality appraisal

### Methods

#### Review Design

This systematic literature review was based on the design synthesis methods of the Evidence for Policy and Practice Information Centre (EPPI-Centre) [30] and the integrative review methodology described by Whittemore and Knaf [31]. Integrative reviews enable the synthesis of data from diverse sources (qualitative and quantitative) to provide a broad and holistic understanding of the subjective and objective elements of a topic, including context, processes, and outcomes [31]. Integrated thematic analysis of data guided a narrative synthesis of the results. Data from qualitative, quantitative, and mixed methods studies were included in this narrative synthesis. The review was registered with PROSPERO (registration number: CRD42019125351).

#### Patient and Public Involvement

In the true spirit of patient and public involvement in research, our team included academics, clinicians, three young people with T1DM, and two of their parents. All team members have contributed to this review, including identifying appropriate search terms, assisting with data extraction and data analysis, and providing comments on various drafts of the manuscript.

#### Search Strategy

We searched PubMed, CINAHL, MEDLINE, Scopus, ProQuest, and Web of Science (search in title/abstract). The search string included the following keywords: (“Type 1 diabetes” OR “insulin dependent diabetes mellitus” OR “juvenile diabetes”) AND (“self manage*” OR “self measure*” OR “self monitor*”) AND (adolescent OR children) AND experience*. We did not use the term “technology” or a similar term in the search string because this limited the results considerably (a comparison was conducted). The reference lists of included studies were searched to include studies that did not appear in the database search. The Cochrane software Covidence [32] was used to assist in the systematic review process from screening to data extraction.

### Inclusion/Exclusion Criteria

Owing to the lack of age specification in many studies, we included studies with participants aged 12 to 25 years to ensure we captured adolescents, who were our primary interest. Studies that focused on parents’ or caregivers’ experiences of caring for a young person with T1DM were also included. We included peer-reviewed studies conducted in any country and in English language from 2009 to early 2019. We excluded randomized controlled trials (RCTs) owing to the integrative narrative scope of the review, which aimed to understand experiences rather than efficacy and effectiveness of technology. Other systematic reviews, conference abstracts, and grey literature were excluded.

### Screening and Quality Assessment

Selected studies were reviewed independently by two researchers, based first on the title and abstract and then on full-text review. Conflicts were resolved through discussion with a third independent reviewer. A full-text quality appraisal

### Methods

#### Review Design

This systematic literature review was based on the design synthesis methods of the Evidence for Policy and Practice Information Centre (EPPI-Centre) [30] and the integrative review methodology described by Whittemore and Knaf [31]. Integrative reviews enable the synthesis of data from diverse sources (qualitative and quantitative) to provide a broad and holistic understanding of the subjective and objective elements of a topic, including context, processes, and outcomes [31]. Integrated thematic analysis of data guided a narrative synthesis of the results. Data from qualitative, quantitative, and mixed methods studies were included in this narrative synthesis. The review was registered with PROSPERO (registration number: CRD42019125351).

#### Patient and Public Involvement

In the true spirit of patient and public involvement in research, our team included academics, clinicians, three young people with T1DM, and two of their parents. All team members have contributed to this review, including identifying appropriate search terms, assisting with data extraction and data analysis, and providing comments on various drafts of the manuscript.

#### Search Strategy

We searched PubMed, CINAHL, MEDLINE, Scopus, ProQuest, and Web of Science (search in title/abstract). The search string included the following keywords: (“Type 1 diabetes” OR “insulin dependent diabetes mellitus” OR “juvenile diabetes”) AND (“self manage*” OR “self measure*” OR “self monitor*”) AND (adolescent OR children) AND experience*. We did not use the term “technology” or a similar term in the search string because this limited the results considerably (a comparison was conducted). The reference lists of included studies were searched to include studies that did not appear in the database search. The Cochrane software Covidence [32] was used to assist in the systematic review process from screening to data extraction.

### Inclusion/Exclusion Criteria

Owing to the lack of age specification in many studies, we included studies with participants aged 12 to 25 years to ensure we captured adolescents, who were our primary interest. Studies that focused on parents’ or caregivers’ experiences of caring for a young person with T1DM were also included. We included peer-reviewed studies conducted in any country and in English language from 2009 to early 2019. We excluded randomized controlled trials (RCTs) owing to the integrative narrative scope of the review, which aimed to understand experiences rather than efficacy and effectiveness of technology. Other systematic reviews, conference abstracts, and grey literature were excluded.

### Screening and Quality Assessment

Selected studies were reviewed independently by two researchers, based first on the title and abstract and then on full-text review. Conflicts were resolved through discussion with a third independent reviewer. A full-text quality appraisal

<table>
<thead>
<tr>
<th>Technology</th>
<th>Acronym</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time continuous glucose monitoring</td>
<td>RT-CGM</td>
<td>This device has a glucose sensor that measures the wearer’s levels of glucose in the interstitial fluid. A signal transmits continuously via radio frequency to a receiver, where the user can see glucose levels in real-time intervals of a few minutes [18,19].</td>
</tr>
<tr>
<td>Continuous subcutaneous insulin infusion</td>
<td>CSII</td>
<td>This form of insulin therapy has been in use for some time. Short-acting insulin is provided through a pump. The dose is adjusted to meet the individual user’s insulin needs, established with experience over time [19].</td>
</tr>
<tr>
<td>Cell phone glucose monitoring</td>
<td>CPGM</td>
<td>This cell phone–based system transmits the user’s blood glucose levels to a host computer, which is monitored by a health care professional [20].</td>
</tr>
<tr>
<td>Flash glucose monitoring</td>
<td>FGM</td>
<td>This device has a sensor that monitors the user’s levels of glucose in interstitial fluid. The user physically swipes a reader device over the sensor to transmit a real-time glucose level and 8 hours of retrospective data, including a trend line [21,22].</td>
</tr>
<tr>
<td>Hybrid closed loop system</td>
<td>HCL</td>
<td>The system is a package comprised of an insulin pump and a CGM system. It can function in the following two different modes: “auto mode” (CL) and “manual mode” (HCL). In CL (auto mode), basal insulin delivery is automatically adjusted in response to CGM levels that are transmitted to the insulin pump. CL is sometimes also called “artificial pancreas” as it requires minimal input from the user. In HCL (manual mode), preprogrammed insulin doses are infused throughout the day, and users must manually deliver bolus doses at meal times and other times to correct blood glucose levels [23,24].</td>
</tr>
<tr>
<td>Multiple dose injection therapy</td>
<td>MDI</td>
<td>This system of insulin delivery has been in use for a long time. It involves subcutaneous injections of either long- or rapid-acting insulin. Long-acting insulin is usually injected once or twice daily and rapid-acting insulin is injected at meal times [25].</td>
</tr>
<tr>
<td>Sensor-augmented pump therapy</td>
<td>SAPT</td>
<td>This system combines CSII and CGM. The glucose sensor is introduced directly into the CSII, and as the name indicates, augments insulin pump therapy [19].</td>
</tr>
</tbody>
</table>

a continuous glucose monitoring.
b closed loop.
c hybrid closed loop.
was performed independently by two reviewers using the Mixed Methods Appraisal Tool (MMAT) [33].

Data Analysis

We combined the study findings in a thematic narrative synthesis. Differences by technologies (CGM, cell phone glucose monitoring [CPGM], FGM, HCL, CL, insulin pumps/bolus advisors, and sensor-augmented pump therapy [SAPT]) were identified within the narrative. Owing to the integrative narrative character of our review, we did not conduct a meta-analysis or report statistical results. This is in line with the narrative synthesis method used in previous systematic reviews [34-36]. We used the quality assessment of the respective studies/papers (MMAT) to ensure credibility of the papers.

Results

Data Extraction and Synthesis

Of 528 identified references, 59 were selected for full-text review. A total of 17 studies were included. Of these, seven studies used qualitative research methods [37-43], four used quantitative methods [20,44-46], and six used mixed method designs [47-52], with only the quantitative component [50] or qualitative component [49,51] of three studies included (Figure 1).

Data were extracted to summarize study characteristics, including study descriptors, technology used, study aims, methods, main findings, and included themes (Multimedia Appendix 1). Data were coded into categories that were classified into eight themes following in-depth discussion and comparison. These themes were representative of common experiences described in the included studies. These provided a structure to systematically examine and discuss the evidence.
Included studies were from the United States (n=7) [20,37,39,41,44,50,52], United Kingdom (n=5) [20,38,47,48,51], Canada (n=2) [42,49], New Zealand (n=1) [40], France (n=1) [46], and Australia (n=1) [45]. Study methodology included in-depth or semistructured face-to-face interviews [38,40,42,43,44,49], surveys and questionnaires [20,44-48,50-52], focus groups [37,49], and analysis of online blog posts and comments [39,41]. Experiences with technologies examined included studies on CGM [38,39,44-49,52], FGM [46], CPGM [20], insulin pump therapy and bolus advisers [43], CSII [45], SAPT [42], and HCL/CL [37,48]. Some studies included experiences of using insulin pumps and/or CGM [40,41,47]. Study sample sizes ranged from 6 to 347, with participants comprised of parents and young people, with ages ranging from 4 to 24 years.

**Quality Assessment**

The consensus rating for all studies on bias was low risk, and thus, none of the 17 studies needed to be excluded because of high risk of bias (Multimedia Appendix 2).

**Thematic Results**

People’s experiences with devices were described within eight themes that included expectations prior to device use on one hand and usage experiences on the other hand. The themes were as follows: (1) expectations of the technology prior to use, (2) impact on sleep and overnight experiences, (3) experiences with alarms, (4) impact on relationships and independence, (5) perceived impact on blood glucose control, (6) device design and features (quality: equipment and size; data and trends: visualization, accuracy, and calibration; invasiveness), (7) cost, and (8) user satisfaction (Multimedia Appendix 3).
Expectations of the Technology Prior to Use

Adolescents expected HCL technology to be self-sufficient, believing it would provide a hands-off experience and live up to its name of an “artificial pancreas,” thereby giving them a break from managing diabetes [37]. Both parents and young people expected that HCL [37], SAPT/CGM/pump [41], and CPGM [20] would reduce the burden of diabetes in their lives. Prior to the use of CL technology, more than half of adolescents and parents reported an expectation of feeling safe when using CL systems, and some parents anticipated that their sleep would be better [48]. However, half of both groups anticipated a negative impact on their usual care routines [48]. At the same time, adolescents worried that CL would draw more attention to their diabetes [48].

Potential users of SAPT expected increased spontaneity and independence, feelings of normality, improved physical performance, and minimized SMBG, as well as reduced hypoglycemic and hyperglycemic episodes in adolescents [42]. Parents expected SAPT to simplify diabetes management and to enable a “normal” life for their child, while adolescents expected that CGM and insulin pump data sharing would reduce parental anxiety at night [40].

Parents believed that SAPT could serve as a second pair of eyes (safety mechanism), especially at night, and that it would help optimize the child’s glycemic control (as measured by HbA1c) to prevent future complications, alleviate stress in the parent-child relationship, and reduce their own anxiety [42]. In general, it was expected that CGM would make life easier for both parents and T1DM children [49], and excitement was expressed about new CGM and pump devices owing to expectations that they might reduce the T1DM management burden [41].

Perceived Impact of Sleep and Overnight Experiences

Seven studies reported results related to overnight device use, including studies on CGM [41,47,49-51], and CL [48] or HCL devices [37,48]. Young participants with T1DM using HCL/CL devices and their parents described waking up feeling better [48], with glucose levels in range [37,48], the benefits of which had an enduring positive effect throughout the day [48]. More stable blood glucose resulted in fewer alarms at night when using CL [48] or HCL [37], and reduced fear of hypoglycemia. Similarly, for (standalone) CGM systems, improved nighttime diabetes management, a feeling of safety and reduced fear, and improved sleep were reported [38,49-51]. Easy access to sensor glucose levels at night increased knowledge [38] and resulted in improved self-management confidence [50].

Some parents in the Health Quality Ontario study [49] reported that despite known long-term risks, before using CGM, they had deliberately kept their child’s blood glucose level high before sleep to avoid overnight hypoglycemic episodes. The use of CGM had enabled better management decisions, including the cessation of this practice. Some parents in this and other studies about CGM stated that the device had saved their child’s life overnight [38,49,51]. Parents also reported disrupted sleep related to CGM due to either false alarms or fear of hypoglycemic events [41,47].

Experiences reported about alarms referred to CGM [38,41,44,47,49,51,52], SAPT [42], and HCL systems [37]. Parents and young people reported a sense of reassurance and safety with CGM alarms, in the knowledge that they provided protection against hypoglycemic episodes [38,49]. Caregivers of children under 18 years of age using CGM found alarms useful in understanding the trending direction of glucose levels [51]. Both CGM [49] and HCL [37] device alarms were considered particularly useful for overnight management. A small number of young people and parents using CGM reported that alarms were the best thing about the device [52]. Users of an HCL system [37] reported fewer overnight interruptions from alarms due to fewer out of range glucose levels.

The benefits of alarms were accompanied by a variety of challenges. HCL users found responding to alarms burdensome [37]. In the Health Quality Ontario study, alarm fatigue amongst adolescents was reported as the most common barrier to the use of CGM [49]. Parents in two studies reported that their children found CGM alarms disruptive during school, which caused some young people to turn them off, impeding optimum diabetes management [38,51]. In one study, parents reported that their children felt nagged by CGM alarms and that they constituted a constant reminder of diabetes in their lives [38]. Interference in daily routine from CGM alarms was reported by more than one-third of participants in a study of young people aged 3 to 25 years [44]. For some parents, alarms were perceived as a sign of their own failure to achieve optimal glycemic control for their child [38].

Both parents and young people reported disrupted sleep related to CGM alarms. In a study of 100 parents of children with T1DM using CGM and insulin pumps [47], the majority of parents reported waking due to the technology, with more than half woken at least four times a week [47], and for one-third of these, the main reason was CGM alarms. Despite CGM alarms, one-fifth of these parents were still fearful of overnight hypoglycemia, and while false alarms were uncommon, they were reported by one-quarter of the parents [47]. Waking due to alarms was reported as frustrating for SAPT users because it was frequently unclear why they went off (whether it was serious or not) [42]. Moreover, alarms went off at inconvenient times and drew attention to the young person, which was perceived as embarrassing [42].

Perceived Impact of Device Use on Relationships and Independence

Eight studies on CL [48], HCL [37], CPGM [20], CGM [38-40,51], and SAPT [42] discussed the impact that devices had on relationships, and nine studies on CPGM [20], HCL [37], CGM [39,40,49,51], SAPT [42], FGM [46], and pump/bolus advisors [43] examined devices and independence of young people in their disease management.

Data sharing oscillated between providing a sense of independence and being a cause of conflict and resentment [39]. On one hand, adolescents and parents felt that SAPT [42], CGM [39,40,49,51], insulin pumps/bolus advisors [43], or CPGM [20] increased the young individual’s independence and
autonomy in managing diabetes as parents did not have to be as hands on as before. This also reduced stress for parents [20] and allowed youth to participate in various leisure activities such as sleepovers, camps, and sports [43,51]. Young people were grateful for the capacity that CGM [40,51] and HCL [37] systems enabled for increased independence and better quality of life, boosting their confidence to try new things and to be more active [40,49,51]. The devices offered freedom to live life in near normality [40,49,51]. Parents also felt that CGM allowed their children to have a sense of safety and of not being alone [39]. Similarly, HCL was reported to result in improved relationships [37] and CL was reported to result in opportunities to talk to people about diabetes (owing to device visibility) [48].

On the other hand, experiences with SAPT included feelings of being tracked and spied on (adolescents) and fear of losing control (parents) [42]. One study that analyzed blogposts from 16 parents of children with T1DM reported that data sharing complicated relationships with a noticeable shift in dependence when adolescents learned to manage their diabetes and parental concerns were perceived as intrusive [39]. In another study about living with SAPT, while some parents reported a desire for their children to use SAPT for “their own peace of mind” [42], they also recognized the negative emotional impact on their child of being accountable for self-management 24 hours a day, and acquiesced to their child’s request to abandon the use of CGM as part of SAPT [42]. These results resulted in some parents and children deliberately refraining from sharing data or at least discussing the boundaries of data sharing [39,42]. Some teenagers preferred to share CGM data with friends they trusted rather than with their parents [39]. In general, parents referred more to partnerships than did young people, approaching management with CGM and insulin pumps as a team, encouraging, and cheerleading, although they were also aware that adolescents often perceived this as nagging [47].

**Perceived Impact on Blood Glucose Levels**

Participants in nine of the included studies reported that using technologies had a positive impact on blood glucose management [20,37,38,44,46-49,51]. steadier blood glucose levels were reported when using HCL [37], and improved blood glucose control was noted with CL [48] and CGM use [44,49,51], with reduced frequency and severity of hypoglycemic events in CGM users [47], as well as lower HbA1c levels when using CPGM [20] and FGM [46]. The majority of caregivers surveyed about the use of both CGM and CSII reported improvements in achieving glycemic targets [47]. Users reported greater confidence and reassurance (CL) [48], and better management decisions (CGM) [49]. Better management also meant less likely over-correction of lows/highs (CGM) [38]. Reduced hypoglycemia-related anxiety was one of the most common perceived benefits of CGM [44]. Overall, parents described CGM as an empowering and motivating tool to fine-tune blood glucose control [38].

**Experiences Related to Device Design and Features**

Participants in 15 studies discussed device design features in terms of device quality [20,38,40-46,48,49,51,52], data characteristics [20,37-42,44,46,48,49,51,52], and discomfort [40,42,44,46,49,51,52].

**Device Quality: Equipment and Size**

One commonly reported disadvantage of CGM [40,44,49,52], SAPT [42], and CL [48] was bulky and heavy sensors and devices. Adolescents experienced challenges with device size and visibility to peers, and described SAPT devices as “ugly” [42]. Managing and wearing additional devices, with increased responsibility, workload, and “hassle,” were reported as parental concerns for CGM [49,51] and SAPT [42], and for young people, it was a constant reminder of living with T1DM [40,49]. In addition, participants did not like the need for CGM backup equipment [40] or second cannulas for CL systems [48].

CGM sensor failures and technical problems, such as sensor cut out and false low values when sleeping on the sensor, were reported [51], in addition to poor FGM [46], HCL [37], and SAPT [42] sensor adhesion (additional tape needed to secure devices) [46] and CGM buttons or power port covers falling off [41]. Children and adolescents had mostly positive experiences with CSII and planned to continue its use as adults [45]. Young people liked that pumps did not require multiple insulin injections [40].

**Data Trends**

Data trends and graphs allowed visualization of changing glucose levels, which made CGM superior to SMBG [38], made understanding CGP trends easier for youth [20], allowed parents to adjust dosage immediately [49], enabled CGM users “to self-correct out-of-range glucose levels” [52], and translated retrospective CGM data analysis into better understanding of diabetes for informed future decisions [38,51]. Yet, constant streaming of CGM data was described as overwhelming at times, and parents and children found that they needed to establish a routine for using the data [39,49,51]. Difficulties interpreting CGM [51] and SAPT [42] data and graphs were also reported. One study of young people’s use of CL reported that parents found greater value in the graphs and trends than did adolescents (CL) [48].

**Data Lag**

Device accuracy and the paradox of inaccurate data due to lag time between the interstitial and capillary blood glucose levels was a key challenge for one-quarter of FGM users [46], with some choosing to discontinue use because of this [46]. The data lag time created a feeling of data distrust for users of CGM [38,51] and SAPT [42], who resorted to SMBG to clarify high and low readings [38,42,51]. Data distrust caused frustration for adolescents who had previously relied on their embodied experiences to understand blood glucose levels but began doubting their decision-making ability [40,42]. Other studies reported that caregivers thought CGM had good data accuracy [41] or that CGM data were accurate [20].

**Connectivity and Calibration**

Parents of young users of CL reported that connectivity and device calibration were the worst aspects of use [48]. Recalibration was perceived as a burden or as frustrating by CGM [38,52], SAPT [42], CL [48], and HCL [37] users. In addition to calibration, users of HCL technology found that the amount of information to be entered about meals, boluses, and corrective insulin dosages was burdensome [37].
Discomfort Related to Devices
Young people reported that the insertion of CGM [38,44,51,52], SAPT [42], and FGM [46] sensors was painful or irritating. For some CGM/pump [38,49] and FGM [46] users, this resulted in reluctance for both future insertion and removal of the sensor, and in discontinued device use [46]. Yet, reduced finger pricking was seen as an advantage of CGM [40,51] and sometimes was the motivation to use new technology (eg, FGM) [46]. Overall, complaints about CGM (including calibration, size, and difficulty inserting the device) were tempered with an emphasis on the benefits users experienced, which they believed outweighed any disadvantages [38,51].

Financial Cost
Four studies from New Zealand [40], Canada [42,49], and the United Kingdom [51] considered the financial cost of SAPT/insulin pumps and CGM devices. Cost issues were cited as the main reason for interrupting or ceasing CGM use in a French study [46] and as a reason for not using CGM in the United States [20]. Parents and adolescents were described as “living worried,” being faced with the stressor of reconciling affordability of SAPT devices with everyday living costs [42]. Parents reported that CGM/SAPT was too expensive to fund themselves owing to the high ongoing supply requirements [42] and the short life span of replaceable sensors [49]. Some used CGM sensors longer than recommended to save money [49]. In Canada, lack of insurance and/or government funding for CGM compared to insulin pumps was cited as a barrier to uptake [42,49]. If asked to choose between an insulin pump and CGM, some parents opted for CGM since they considered continuous data and information more valuable than the flexibility offered by a pump [49].

Satisfaction With the Technology
One US study of 208 youth aged 8 to 18 years and their parents [52] measured satisfaction using the Continuous Glucose Monitoring Satisfaction Scale (CGM-SAT), which includes 5-point Likert subscales on the “benefits of CGM” and “hassles of CGM.” Parents’ and adolescents’ responses were compared, as was CGM use in terms of days per week. Overall, satisfaction with CGM technology was higher for parents compared to young people [52]. Frequent users who used CGM for over 6 days per week reported considerable higher satisfaction compared with those who used CGM for less than 4 days per week [52]. In another US study, among 35 families using the mySentry CGM system [50], parents reported high levels of satisfaction with overnight monitoring of their child’s glucose levels. In a French study of 347 FGM users aged 0 to 18 years, overall satisfaction was high, with two-thirds of users reporting being satisfied [46]. The most frequent motive for dissatisfaction with FGM was the absence of real-time alerts [46]. Regarding CL technology, overall, there were favorable responses in terms of impact and satisfaction [48].

Discussion
Principal Findings
The eight themes that emerged from our review of the 17 included studies illustrate the impacts of diabetes and the associated use of technology on various aspects of young people’s and their caregivers’ lives.

Our results showed that expectations prior to technology use could be split into expectations that could not be met with the current state of the technology (eg, artificial pancreas [37]) and expectations that were pretty much mirrored by the reported experiences (eg, improved safety). Experiences partly depended on the particular technology used. The majority of the papers focused on CGM and/or insulin pumps, with some reporting experiences specific to the respective devices (eg, CGM sensor accuracy/failure). However, as the results for CGM and insulin pumps are frequently reported together, further research is needed to examine if the difference in the devices is a key factor in user experiences.

Sleep disturbances due to alarms in youth and caregivers, and overnight management have been reported as major challenges in T1DM management in previous research [53], along with anxiety and fear of hypoglycemia in both youth and their caregivers [54]. Efficient and reliable hypoglycemia alert systems that do not disrupt sleep to an extent that affects overall management still have to be developed.

While parents are solely responsible for disease management of young children, the dynamics of care coordination change in adolescence, requiring fine balancing of parental support and involvement [11]. Adolescence is a time when children seek to achieve increasing independence and to separate emotionally from their parents, prioritizing relationships with their age peers. During this time, diabetes can impact the many important relationships of young people, including relationships with their parents, health professionals, teachers, and peers [20]. Our results indicate that automatized monitoring systems and insulin pumps offer potential for greater independence in adolescents and reduce the ongoing monitoring and management burden for parents [55]. At the same time, technologies can negatively affect the relationship between adolescents and their caregivers (eg, data sharing complicates relationships). Young people’s expectations of technology often diverge from those of their caregivers, and priorities are set differently (eg, independence versus reduced fear of hypoglycemia and improved sleep). Moreover, stigmatization [56] and judgement [57] by family members or peers can affect relationships and overall diabetes management. Thus, the nature of relationships between young people with T1DM and their caregivers, peers, and health professionals needs to be accounted for in the design of these technologies, particularly the relationship between youth with T1DM and their parents, which is characterized by a fine balance between autonomy and dependence (interdependence, also termed as transactional) [58]. Reliable devices are needed to engender trust and encourage practices that optimize diabetes management, avoiding risky behaviors that were reported by some participants in this review (eg, parents allowing higher than desirable blood glucose levels to avoid overnight hypoglycemia) [59].

Diabetes technology has been shown to be effective in improving metabolic control [6] in young people with T1DM at an early stage of the disease, preventing long-term complications (referred to as “metabolic memory”) [60]. Similar
Successful diabetes technology use and improved self-care, which are reflected in improved blood glucose levels, can be achieved when individual empowerment is promoted [64,65]. Thus, a particular focus should be put on empowerment practices when designing diabetes technology for self-management. This can be achieved through user-centric design, which can aid in removing barriers to use at the same time, enabling the development of systems that are suitable for long-term use [66].

User expectations and preferences in technology design need to be accounted for (eg, reduction in device size and improved device quality as mentioned in our review).

Cost and funding issues hindered technology uptake and potential T1DM self-care in the included studies. While government subsidies are available for blood glucose meters in New Zealand, users in our review reported frequent changes by the government, which forced them to acquire newer and cheaper devices more prone to inaccurate measurements. Lack of insurance and/or government funding for CGM systems in Canada and the United Kingdom, and for CPGM systems in the United States [20] has been reported as an uptake barrier in the studies included in our review. FGM became reimbursable in France under the French National Health Insurance program in 2017 [46]. In Australia, subsidized schemes of CGM for children and adolescents have been expanded by the government to include FGM starting from 2020, but for many, these schemes cut out at the age of 21 years [67]. This shows that funding for new diabetes technology varies widely among countries, impacting technology uptake and use.

Despite a variety of reported challenges in using technologies to manage T1DM, overall, the studies in our review examining satisfaction with use reported high levels of satisfaction, and benefits were predominant. This is congruent with previous research that found new technology use is frequently accompanied with increased satisfaction with the technology when compared to multiple dose injections and SMBG [68].

Owing to its perceived benefits, there is a growing desire among the young T1DM community for automated CL “artificial pancreas systems” that integrate CGM with insulin delivery [69]. Yet, these expectations and desires are frequently not met in actual experiences with available technology. Even though available systems are a step toward automation of diabetes control, our review demonstrates that current technology is insufficient to provide fully reliable and sustainable automated systems that fulfill the expectations of young people with diabetes and their caregivers. The gap between “ideal” device systems, such as CL systems (artificial pancreas), and the currently available status quo of systems (eg, sensors and HCL systems) is a barrier that warrants further development. There is a need for improved and advanced diabetes technologies complying with the various user requirements outlined above.

The strength of this review lies in its unique focus on young individuals with T1DM, as this population is among those that experience what has been identified as “diabetic distress” and that undergo the most difficulty in adapting to diabetes needs and are most challenged in terms of glycemic variability [63].

Implications for Practice

The conglomeration of experiences and attitudes associated with currently available diabetes devices and technologies is a step toward a possible refinement of future diabetes technologies. Our review supports a move toward a tailored approach for individuals with T1DM to create technology that is robust, intuitive, and sustainable. An integrative approach involving adolescents, parents, health care providers, and teachers should be used to develop future technology and guide design experiments. Individuals with T1DM from diverse ethnic and socioeconomic backgrounds also need to be included in the co-design process to advance T1DM technology. This includes discussions of use and sharing of data. Our review has shown that while access to continuous data was valued by CGM users, there were also challenges in managing the amount of data. This resonates with a clinical evidence review of 22 studies that found that data could be perceived as overwhelming for some users [49]. Challenges like these must be addressed in collaboration with young people with T1DM and their caregivers.

Study Limitations

While our main interest was in examining adolescents’ and their caregivers’ experiences of using devices, some included studies also involved younger children and older youth. It was not possible to exclude these data from our analysis, and at times, these have been included in our analysis.

We did not examine the grey literature, and thus, we might have excluded reports and evaluations that also included experiential data. We only examined studies reported in English, which excludes analysis of experiences in non–English-speaking countries and perhaps young non–English-speaking people’s experiences of using devices in English-speaking countries.

Owing to the rapid evolution of technology and associated changes regarding available devices and systems, there are challenges in evaluating a large number of experiences with a particular device.

Conclusion

Overall, the use of diabetes technology was found to be beneficial and to positively impact disease management for both young people and their caregivers. The included studies reported the advantages of diabetes technologies, such as improved self-management and diabetes outcomes, in young people associated with improved monitoring, data tracking, and data sharing, as well as decreased anxiety and psychological pressure in both parents and children. However, technology did not always live up to users’ expectations. Several barriers and challenges toward its use were reported, such as cost, the size...
and visibility of devices, and the intrusiveness of alarms, which drew attention to the fact that the user had T1DM. Continued use of diabetes technology was underpinned by its benefits outweighing its challenges, especially among younger people. Collaboration with young people and their caregivers is essential to ensure that future T1DM technologies meet their expectations and needs.

Acknowledgments
This research was funded by and has been delivered in partnership with Our Health in Our Hands (OHIOH), a strategic initiative of the Australian National University, which aims to transform health care by developing new personalized health technologies and solutions in collaboration with patients, clinicians, and health care providers. AT gratefully acknowledges the support of the Australian Research Council (ARC) (DP190101864 and FT200100939) and NATO Science for Peace and Security Program.

Authors’ Contributions
MC, NBS, AP, and JD had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors were involved in study concept and design. MC and JD acquired the data and conducted the initial analysis. All authors were involved in the subsequent analysis and interpretation of the data. MC, NBS, AP, and JD were involved in drafting the manuscript; all authors were involved in revision. JD supervised the study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Data extraction table of included studies.
[DOCX File, 34 KB - diabetes_v611e20973_app1.docx ]

Multimedia Appendix 2
Quality assessment using the Mixed Methods Appraisal Tool (MMAT).
[DOCX File, 34 KB - diabetes_v611e20973_app2.docx ]

Multimedia Appendix 3
Themes derived from included studies.
[DOCX File, 31 KB - diabetes_v611e20973_app3.docx ]

References


67. We’re for more choice to help Australians living with diabetes – it is our #1 priority. Abbott Diabetes Care. 2020 Feb 02. URL: https://www.freestylelibre.com.au/ndss [accessed 2021-01-18]


Abbreviations

- CGM: continuous glucose monitoring
- CL: closed loop
- CPGM: cell phone glucose monitoring
- CSII: continuous subcutaneous insulin infusion
- FGMM: flash glucose monitoring
- HCL: hybrid closed loop
- SAPT: sensor-augmented pump therapy
- SMBG: self-monitoring of blood glucose
- T1DM: type 1 diabetes mellitus

http://diabetes.jmir.org/2021/1/e20973/ JMIR Diabetes 2021 | vol. 6 | iss. 1 | e20973 | p.134 (page number not for citation purposes)
Public Perspectives on Anti-Diabetic Drugs: Exploratory Analysis of Twitter Posts

Su Golder¹, BSc, MSc, PhD; Millie Bach¹, MSc; Karen O’Connor², MSc; Robert Gross³, MD, MSCE; Sean Hennessy², PharmD, PhD; Graciela Gonzalez Hernandez², PhD

¹Department of Health Sciences, University of York, York, United Kingdom
²Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Pennsylvania, PA, United States
³Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Pennsylvania, PA, United States

Corresponding Author:
Su Golder, BSc, MSc, PhD
Department of Health Sciences
University of York
Heslington
York, YO10 5DD
United Kingdom
Phone: 44 01904321904
Email: su.golder@york.ac.uk

Abstract

Background: Diabetes mellitus is a major global public health issue where self-management is critical to reducing disease burden. Social media has been a powerful tool to understand public perceptions. Public perception of the drugs used for the treatment of diabetes may be useful for orienting interventions to increase adherence.

Objective: The aim of this study was to explore the public perceptions of anti-diabetic drugs through the analysis of health-related tweets mentioning such medications.

Methods: This study uses an infoveillance social listening approach to monitor public discourse using Twitter data. We coded 4000 tweets from January 1, 2019 to October 1, 2019 containing key terms related to anti-diabetic drugs by using qualitative content analysis. Tweets were coded for whether they were truly about an anti-diabetic drug and whether they were health-related. Health-related tweets were further coded based on who was tweeting, which anti-diabetic drug was being tweeted about, and the content discussed in the tweet. The main outcome of the analysis was the themes identified by analyzing the content of health-related tweets on anti-diabetic drugs.

Results: We identified 1664 health-related tweets on 33 anti-diabetic drugs. A quarter (415/1664) of the tweets were confirmed to have been from people with diabetes, 17.9% (298/1664) from people posting about someone else, and 2.7% (45/1664) from health care professionals. However, the role of the tweeter was unidentifiable in two-thirds of the tweets. We identified 13 themes, with the health consequences of the cost of anti-diabetic drugs being the most extensively discussed, followed by the efficacy and availability. We also identified issues that patients may conceal from health care professionals, such as purchasing medications from unofficial sources.

Conclusions: This study uses an infoveillance approach using Twitter data to explore public perceptions related to anti-diabetic drugs. This analysis gives an insight into the real-life issues that an individual faces when taking anti-diabetic drugs, and such findings may be incorporated into health policies to improve compliance and efficacy. This study suggests that there is a fear of not having access to anti-diabetic drugs due to cost or physical availability and highlights the impact of the sacrifices made to access anti-diabetic drugs. Along with screening for diabetes-related health issues, health care professionals should also ask their patients about any non–health-related concerns regarding their anti-diabetic drugs. The positive tweets about dietary changes indicate that people with type 2 diabetes may be more open to self-management than what the health care professionals believe.

(JMIR Diabetes 2021;6(1):e24681) doi:10.2196/24681

KEYWORDS
diabetes; insulin; Twitter; social media; infodemiology; infoveillance; social listening; cost; rationing
Introduction

In 2016, 4.2 million diabetes-related deaths were reported worldwide [1], which makes diabetes the seventh leading cause of mortality [2]. For both type 1 and type 2 diabetes, treatment and management aim to achieve adequate glycemic control [3]. Medication nonadherence is reported to be high for insulin and even higher for noninsulin anti-diabetic drugs [4,5]. Patients’ beliefs about medications, such as whether they are perceived to be essential or whether they have side effects, can influence both adherence and self-management behaviors [6]. The odds of nonadherence is 3.4 times as high in those who believe that anti-diabetic drugs have serious side effects and 14.3 times as high in people who believe that diabetes treatment regimens are too complex [7].

Given social media’s ability to connect large numbers of people and thereby generate large volumes of data, it has become a novel area for health research and a powerful tool to understand public perceptions. This study uses a particular social media site, that is, Twitter. As a popular social media outlet, Twitter is both a microblogging site and a social networking platform [8]. Since its conception in 2006 [9], Twitter’s popularity has grown to a reported 330 million monthly active users in 2019 [10]. The utilization of Twitter as a data collection platform is increasing and it is the most commonly utilized social media platform within health research [11]. Sinnenberg et al [12] demonstrated that the number of health-related studies harnessing Twitter in 2015 was over 10 times higher than that in 2010, and their systematic review of 137 studies identified many ways in which Twitter data can be used. The most common Twitter analyses identified by the authors were content analyses, wherein the words, pictures, or sentiment of tweets are analyzed. The monitoring of vocabulary within tweets for pharmacovigilance purposes is an expanding area of research [13], while the exploration of tweets discussing perceptions of medications can help understand compliance and therapeutic decision making [14]. With regard to diabetes, studies have examined changing sentiments in Tweets on diabetes since the COVID-19 outbreak [15], and public perceptions have been examined on Twitter in detail for diseases such as Ebola virus disease [16] and cancer [17] and products such as e-cigarettes [18].

In this study, we sought to identify perceptions held by people discussing anti-diabetic drugs on Twitter. In particular, we sought to assess 3 questions: (1) Who discusses anti-diabetic drugs on Twitter? (2) Which anti-diabetic drugs are the most frequently discussed on Twitter? and (3) What are the most common health-related topics discussed on Twitter regarding anti-diabetic drugs?

Methods

Publicly available tweets posted between January 1, 2019 and October 1, 2019 were retrieved by the University of Pennsylvania’s Health Language Processing Center [19] from a large publicly available data set curated by the Internet Archive. The Internet Archive is a nonprofit organization that builds digital libraries of internet sites and provides free access to the data to researchers. We removed retweets from the collection. We selected this time scale in order to account for any seasonal or newsworthy variations in the tweets posted. Search terms associated with anti-diabetic drugs, including generic names, brand names, and common misspellings (Multimedia Appendix 1) were used to retrieve 10,308 tweets (Figure 1). After removing 515 duplicates, 92.9% (9107/9793) of the medication-related tweets were found to be about insulin.

We, therefore, constructed a purposive sample of all tweets about noninsulin anti-diabetic drugs (n=686) so as to not lose any potential valuable information and a random sample about insulin (n=3314).

Qualitative studies traditionally have small sample sizes [20], but social media analyses are associated with qualitative data on a quantitative scale [21]. Consequently, qualitative Twitter analyses often use a sample of tweets rather than the full sampling frame [22]; sample sizes range from a few hundred [23] to thousands of tweets [12]. Guided by previous research, we initially began with 4000 random tweets (4000/9793 or 40.8% of our total sample), with additional samples to be analyzed if code saturation and meaning saturation were not met. Code saturation can be defined as the point at which all codes have been identified, while meaning saturation is the point at which all codes are understood [24]. After coding all 4000 tweets, code saturation and meaning saturation appeared to have been met [24] and a further sample was not necessary.

Codes are labels for assigning units of meaning [25]. In qualitative content analysis, the use of codes results in the generation of themes that can be used to interpret the meaning of the text [26]. Health-related tweets were coded based on the perception expressed in the tweet. This used the conventional content analysis inductive framework proposed by Hsieh and Shannon [27] to explore both the manifest and latent meanings of the tweets and ensured that the codes arose from the data itself rather than being predefined. An inductive approach was particularly useful as there is little theory on anti-diabetic drug perceptions discussed via Twitter on which to base any assumptions and there is no particular framework to work from. Inductive approaches on Twitter data are also commonplace in the scientific literature [16]. Initial codes were given to each tweet, and upon reflection of the whole data set, similar or linked codes were clustered into themes. Some similar themes were further combined to form subthemes under an overarching theme.
The themes identified at this stage formed the basis of the coding scheme. We created a manual containing the coding scheme and instructions with examples on how to correctly assign codes. We filtered the Internet Archive data set by matching the keywords list, which includes all anti-diabetic drugs and their variants in the tweets. Only tweets in English and those that were not retweets were retrieved. The output file created contains all tweets where a match was found and included the user ID, tweet ID, tweet text, data created, and the keyword that matched in separate columns in an Excel. The keyword column helped ascertain the drug mention; however, the themes were hand-coded from scratch [28].

Two researchers independently coded 231 tweets by using the coding scheme. A random sample of 231 tweets was found to be sufficient to measure agreement and to stimulate discussion on the coding scheme as all codes were represented multiple times in this sample size. Because the initial kappa coefficient was 0.67, disagreements were discussed, and the coding instructions adapted accordingly. A further 169 tweets were then coded independently by 2 reviewers, producing a satisfactory kappa score of 0.73 [29]. Each of the remaining tweets was then coded by one of the two researchers, with all codes checked by the other reviewer and any disagreements resolved by discussion. First, tweets were coded for whether they truly were anti-diabetic drug–related. Second, any anti-diabetic drug–related tweets were coded as either health-related or non–health-related. Health-related tweets were further coded. Tweeters were categorized as (1) those who used the drug themselves, (2) people who knew someone who takes the drug, (3) health care providers, or (4) unclear, that is, the relationship between the tweeter and the anti-diabetic drug was unclear. Figure 2 shows a theoretical tweet, which has been coded, to show how coding was performed.

The availability of social media data means that it is relatively easy to trace quotations back to the user; therefore, there is a risk of deductive disclosure [30]. This makes reporting direct quotations problematic. Subtle changes to tweets are at odds with the Twitter display requirements, which prevent the alteration of tweets [31]. We, therefore, undertook a descriptive approach through paraphrasing tweets and by only directly quoting commonly used terms so that they cannot be traced back to an individual tweet. All data used in this study were collected according to the Twitter terms of use and were publicly available at the time of collection and analysis. We have an institutional review board certificate of exemption from the University of Pennsylvania. Each theme was explored regardless of how often it occurred.
Results

Tweeter Description

The results of this study are based on the 1664 health-related tweets (Table 1). A quarter (415/1664, 24.9%) of the tweets were by patients with diabetes taking anti-diabetic drugs, or who had taken the anti-diabetic drug in the past or who might initiate the anti-diabetic drug in the future: 87 (21.1%) of these self-identified as having type 1 diabetes, 61 (14.6%) as having type 2 diabetes, 2 (0.5%) as having gestational diabetes, and 2 (0.5%) as having secondary diabetes. The type of diabetes could not be classified for two-thirds of the tweeters; 17.9% (298/1664) of the tweets were second-person accounts, often about a family member or a person in a news story, and 2.7% (45/1664) of the tweets were from health care professionals. We could not establish the relationship between the tweeter and the anti-diabetic drug for the remaining 54.4% (906/1664) of the tweets.

Table 1. Proportions of the types of tweets and tweeters.

<table>
<thead>
<tr>
<th>Type of tweet/type of tweeter</th>
<th>n (%)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Irrelevant tweets (n=2336)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non–health-related</td>
<td>1556</td>
<td>(66.6)^a</td>
</tr>
<tr>
<td>Tweets that mention an anti-diabetic drug but are not directly related to health, for example, jokes, advertisements.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a drug</td>
<td>693</td>
<td>(29.6)</td>
</tr>
<tr>
<td>Key term is used but is not in reference to a drug, for example, using the term “insulin” to mean the endogenous hormone rather than the exogenous anti-diabetic drug.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in English</td>
<td>7</td>
<td>(0.3)</td>
</tr>
<tr>
<td>The majority of the tweets were not in English.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not related to diabetes</td>
<td>80</td>
<td>(3.4)</td>
</tr>
<tr>
<td>Tweet refers to drug being used for a purpose other than diabetes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health-related tweets (n=1664)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-person report</td>
<td>415</td>
<td>(24.9)</td>
</tr>
<tr>
<td>Tweet from a diabetic person—uses phrases like “my drug…”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second-person report</td>
<td>298</td>
<td>(17.9)</td>
</tr>
<tr>
<td>Tweets from someone who is not diabetic but is about a diabetic person—uses phrases like “my daughter’s drug…”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care professional</td>
<td>45</td>
<td>(2.7)</td>
</tr>
<tr>
<td>Tweet is from a health care professional—uses phrases like “my patient’s drug”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>906</td>
<td>(54.4)</td>
</tr>
<tr>
<td>There is insufficient context to determine who is sending the tweet.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^aOf these, 920 (59.1%) tweets were on cost.

Anti-Diabetic Drugs Under Discussion

Tweets related to 33 anti-diabetic drugs across 11 drug classes were identified: insulin (1281 tweets), biguanides (194), SGLT2 inhibitors (102), DDP4 inhibitors (33), GLP1 agonists (97), sulfonylureas (11), thiazolidinediones (16), metformin (2), α-glucosidase inhibitors (1), meglitinides (1), and amylase analogues. People tweeted using both generic and brand names.

Common Perceptions

We identified 13 themes (Table 2). In most cases, we could not determine if the tweet was about type 1 or type 2 diabetes. Cost and efficacy dominated type 1 diabetes posts and other treatments, and adverse drug reactions dominated type 2 diabetes tweets. Type 1 diabetes tweets were also more likely to discuss more than one topic (Figure 3).
Table 2. Themes of the health-related tweet categories (n=1664).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Explanation</th>
<th>Subthemes</th>
<th>n (%), Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Tweet discusses the cost of an anti-diabetic drug in relation to health issues.</td>
<td>How much do anti-diabetic drugs cost? Attitudes toward cost, insurance problems, health consequences, social consequences, managing cost</td>
<td>669 (40.2)</td>
</tr>
<tr>
<td>Efficacy</td>
<td>Tweet discusses efficacy of the drug, both positive and negative. This includes tweets about the necessity of the drug and tweets that state that death will occur if the anti-diabetic drug is not taken.</td>
<td>Positive and negative</td>
<td>465 (27.9)</td>
</tr>
<tr>
<td>Information resource</td>
<td>Tweet provides information about the anti-diabetic drugs. These tweets reference research articles or clinical guidelines rather than someone’s belief about the anti-diabetic drugs.</td>
<td>Links and information summaries</td>
<td>371 (22.2)</td>
</tr>
<tr>
<td>Availability</td>
<td>Tweet discusses the availability of or access to anti-diabetic drugs.</td>
<td>Nationwide availability, personal availability, ensuring availability</td>
<td>158 (9.5)</td>
</tr>
<tr>
<td>Nonadherence</td>
<td>Tweet discusses someone not following the recommendation for taking the anti-diabetic drugs.</td>
<td>Taking too much, taking too little, consequences of nonadherence</td>
<td>124 (7.5)</td>
</tr>
<tr>
<td>Personal opinion</td>
<td>Tweet discusses a personal belief about anti-diabetic drugs.</td>
<td>Preferences, opinions of people without diabetes, opinions of people with diabetes</td>
<td>94 (5.6)</td>
</tr>
<tr>
<td>Other treatment options</td>
<td>Tweet compares an anti-diabetic drug to another management option for diabetes.</td>
<td>Other management options, effect on anti-diabetic drug, attitudes toward other treatments</td>
<td>54 (3.2)</td>
</tr>
<tr>
<td>Question</td>
<td>Tweet is being used to seek advice or to challenge others.</td>
<td>Advice from others, educational tool</td>
<td>41 (2.5)</td>
</tr>
<tr>
<td>Changes to treatment</td>
<td>Tweet discusses starting, stopping, or changing to another anti-diabetic drug.</td>
<td>Starting a medication, stopping a medication, changing insulin delivery</td>
<td>31 (1.8)</td>
</tr>
<tr>
<td>Stigma</td>
<td>Tweet discusses stigma surrounding anti-diabetic drugs.</td>
<td>Specific situations associated with insulin delivery, reducing stigma, opinions of people without diabetes</td>
<td>29 (1.7)</td>
</tr>
<tr>
<td>Dose</td>
<td>Tweet discusses dosing of anti-diabetic drugs. This includes stating the dose, saying how it is taken, or general statements about having to change the dose.</td>
<td>Stating the dose and calculating doses</td>
<td>28 (1.6)</td>
</tr>
<tr>
<td>Adverse drug reaction</td>
<td>Tweet is about an experience of an adverse drug reaction. These should be tweets about adverse drug reactions that have actually happened, rather than beliefs about the potential side effects of an anti-diabetic drug.</td>
<td>Specific side effects, general side effects, associated with insulin delivery</td>
<td>21 (1.3)</td>
</tr>
<tr>
<td>Abuse</td>
<td>Tweet discusses taking the anti-diabetic drug for nonmedical reasons.</td>
<td>Intent to kill or for fun</td>
<td>10 (0.6)</td>
</tr>
<tr>
<td>Nonclassifiable</td>
<td>Some tweets did not provide enough context to determine what it was about.</td>
<td>Too short or incomprehensible</td>
<td>85 (5.1)</td>
</tr>
</tbody>
</table>
Figure 3. Tweet categories by people with type 1 and type 2 diabetes. ADR: adverse drug reaction.

Anti-Diabetic Drugs Are Too Expensive

The cost of insulin was the most common topic. Some tweeters listed the cost while others described it as “too expensive” (669/1664, 40.2%). Tweeters also remarked that the cost had “skyrocketed.” Health care practitioners were aware that the high cost affected the health of their patients. They described how prices had increased during their time and how they tried to prescribe low-cost anti-diabetic drugs. Cost was an issue for both those with and without health insurance coverage. Certain insurance plans cover certain drugs but not insulin. Younger people expressed fears about aging out of their parents’ insurance.

It was generally felt that high costs were unfair and the profit margin too great. Many believed that anti-diabetic drugs should be free. This was fueled by comparisons of the costs outside the United States or comparisons to other medications. The health consequences of being unable to afford anti-diabetic drugs were extensively discussed. Tweeters expressed difficulty in achieving blood glucose level targets, which they reported resulted in long-term repercussions such as losing limbs, going blind, renal failure, and strokes. Diabetic ketoacidosis was mentioned as a specific concern, and the worst case scenario was death. There were also economic and social consequences such as bankruptcy and homelessness. Some tweeters had made lifestyle decisions based solely on their need for anti-diabetic drugs such as taking a job with insurance rather than a preferred job. Tweeters were open in discussing ways of affording anti-diabetic drugs, including asking other tweeters for money, selling their belongings, or working more than one job. Alternative options were buying cheaper anti-diabetic drugs from abroad, buying over-the-counter medicines, or turning to the black market. Large-scale approaches to making anti-diabetic drugs more affordable included using Twitter to promote campaigns such as the #InsulinForAll movement (a campaign launched in the lead up to World Diabetes Day in 2014 by The Pendsey Trust and T1 International) and to contact people in power, with tweets being sent to the US President and pharmaceutical companies.
Anti-Diabetic Drugs Have Varying Efficacy
There was an agreement that insulin was lifesaving. Short-term benefits such as glucose control were noted, as well as generally feeling better. Some tweeters reported issues with their insulin such as insufficient blood glucose reductions, and there were concerns about “Walmart insulin,” with some posts claiming that it is ineffective and caused hypoglycemia. Noninsulin anti-diabetic drugs were perceived to have different levels of efficacy (465/1664, 27.9%). For instance, exenatide and empagliflozin were viewed as effective in reducing weight, which was viewed favorably. Another SGLT2 inhibitor, canagliflozin, was reported to prevent microvascular complications. Metformin had mixed reviews; some felt it worked while others did not.

Wealth of Information on Anti-Diabetic Drugs
Information was mostly tweeted as links to or summaries of journal articles (371/1664, 22.2%). Articles varied from laboratory studies to efficacy evaluations. Studies exploring alternative methods of insulin delivery and the use of noninsulin anti-diabetic drugs as adjunct therapies in type 1 diabetes were considered particularly important. Information also came in the form of videos and links to reports on drug approvals and safety published by regulatory bodies.

Anti-Diabetic Drugs Are Not Always Available
Problems in availability included delays in mail orders, stolen, or lost medication (158/1664, 9.5%). There were posts calling for wider availability of nonprescription insulin. Some tweeters reported use of nonofficial outlets, and Twitter was used to find, sell, or give away extra supplies. Others discussed anti-diabetic drug availability on a national scale. The main topic concerning the United Kingdom was the impact of leaving the European Union. Additional barriers in the United States were the government shutdown from December 22, 2018 to January 25, 2019 [32], which caused financial and logistic issues, impaired access for deported immigrants, and US sanctions on Venezuela. Tweeters were proactive in discussing ways to ensure their anti-diabetic drug supply, such as stockpiling in the United Kingdom or traveling to Canada or Mexico from the United States. However, there were concerns over stockpiling due to storage issues and insulin’s shelf-life and a strong sense that people should not need to travel abroad to receive life-saving medications.

Adherence Can Be Difficult
The majority of tweeters reporting nonadherence mentioned missing doses (124/1664, 7.5%). Those mentioning metformin or lixibliflumate simply stated they had missed a dose, while insulin users provided more detailed reasons. Some forgot to take their insulin or had equipment problems; others deliberately choose not to take it. Reasons for this included dislike of needles, reactions to news stories condemning insulin, diabulimia with tweeters restricting their insulin intake to control their weight, and incorrectly following advice (this included injecting insulin through clothes or failing to take bolus insulin if not eating due to illness). The most commonly cited reason for nonadherence was cost (85/124, 68.5%), which led to rationing either by taking less insulin per injection or by omitting injections. Some who were not then rationing expressed fears about having to do in the future. Insulin overdoses were less commonly discussed, with causes including misreading the dose volume or accidentally taking 2 injections. The only issue reported by tweeters who took an overdose was hypoglycemia.

Tweeters Hold a Range of Personal Beliefs
Some Tweeters stated preferences for particular anti-diabetic drugs that had no scientific evidence for the mechanism of action (94/1664, 5.6%). For instance, there was a perception that insulin makes type 2 diabetes worse. Tweeters with diabetes were mostly negative about being on anti-diabetic drugs, expressing that anti-diabetic drugs make life difficult. Some of these negative attitudes centered around equipment, including not liking the “huge” exenatide needles or the hassle of changing cartridges in prefilled insulin pens.

Anti-Diabetic Drugs Are Considered Alongside Other Treatments
Anti-diabetic drugs were discussed alongside lifestyle changes, particularly diet changes and specific diets, including the ketogenic diet or a vegan lifestyle (54/1664, 3.2%). Mentions of herbal treatments centered around a news story about the death of a person with type 1 diabetes whose herbalist advised the person to stop his/her insulin. Those using alternative or supplementary treatments were happy to do so, and many expressed annoyance at being offered anti-diabetic drugs with no option of management through lifestyle changes. Subsequently, these alternative treatments were discovered through social media or personal research rather than being initiated by a health care provider. The only alternative treatments that health care providers tweeted support for were exercise and ketogenic diets. Those with type 1 diabetes expressed frustration at being told to try nondrug treatments, particularly diet changes. Although they recognized that reducing carbohydrate intake can reduce insulin requirements, some felt the need to state that type 1 diabetes requires insulin, regardless of diet.

Anti-Diabetic Drugs Generate Questions
Those struggling to adjust their anti-diabetic drugs to adequately control their blood glucose levels sought advice from others, and there were questions about where to source “cheap” insulin (41/1664, 2.5%). Health care professionals asked their peers questions, including on the correct anti-diabetic drug, on theoretic scenarios, or interpretation of study findings.

Anti-Diabetic Drug Regimens Can Change
Tweeters with type 2 diabetes actively tried to avoid starting insulin. Similarly, stopping insulin was seen as an achievement. Those who had previously managed with only lifestyle changes felt apprehensive about initiating medications. Some tweeters completely stopped their anti-diabetic drugs, usually with guidance from health care providers and changing to a nondrug therapy. Insulin users reported changing to different types of insulin or administration method rather than a different class of anti-diabetic drugs. These data were captured from 1.8% (31/1664) of the tweets.


**Anti-Diabetic Drugs Are Associated With Stigma**

Taking insulin injections in the public resulted in perceptions of being judged or objection to the practice. Those wearing an insulin device or with scars and bruising due to needles felt these drew unwanted attention. Stigma was greater at airport checkpoints, work, or school. These data were captured from 1.7% of the tweets (29/1664). Some tweets discussed a reduction in stigma. This included restaurants providing carbohydrate content information to facilitate insulin dosing and the sense of togetherness when an individual saw other patients with diabetes taking injections. Some tweeters who did not have diabetes believed that there was no stigma for patients with diabetes, arguing that, “patients with diabetes are not judged for using insulin; so, why should people with depression be judged for taking antidepressants?”

**Dosing Varies Based on the Anti-Diabetic Drug**

Dosing based on meal-time carbohydrate or protein intake was noted to be difficult. Some tweeters shared their calculations. Some tweeters admitted to guessing their doses but that was not effective. For tweeters on non-insulin anti-diabetic drugs, doses were decided upon by health care providers. These data were captured from 1.7% of the tweets (28/1664).

**Anti-Diabetic Drugs Can Cause Adverse Drug Reactions**

The explicitness of the descriptions of the adverse drug reactions varied. Gastrointestinal issues, including vomiting or stomach aches, were mentioned for metformin and empagliflozin. Insulin and pioglitazone were both reported to cause weight issues. Other adverse drug reactions included allergic reactions to insulin, cognitive issues with metformin, and blood count changes with empagliflozin. Some adverse reactions were specific to the mode of insulin delivery, including local skin reactions to injections and scar tissue formation following the use of pumps. Other tweeters stated they had an adverse reaction but did not explain further. Tweeters discussed ways to cope, such as by spreading out the doses. The only adverse reaction that seemed to cause cessation was near-death experiences in 3 cases. These data were captured from 1.6% of the tweets (28/1664).

**Anti-Diabetic Drugs Can Be Abused**

There were first-person reports of deliberately taking too much insulin for the thrill of trying to restabilize blood glucose levels. Insulin was recognized as potentially deadly—there were tweets about people trying to kill themselves or someone else by administering insulin. These data were captured from 0.6% of the tweets (10/1664).

**Non–Health-Related Tweets**

While this study’s primary focus was the exploration of health-related tweets, it became evident that trends within the non–health-related tweets were also important (1556/1664). Though some non–health-related tweets were jokes or advertisements, 59.1% (920/1556) of the tweets were on the cost of anti-diabetic drugs—these raised similar issues to the health-related cost tweets without discussing the health implications.

**Discussion**

**Overview**

This study explored public perceptions of anti-diabetic drugs via the analysis of health-related tweets. We found that the issue of cost dominated both health and non–health-related tweets regarding insulin and overwhelmed our results, with implications for other identified themes such as availability, adherence (via rationing), and safety of cheaper versions. We found a similar proportion of health-related tweets in our sample (1664/4000, 41.6%) when compared to that in our study on statins (5201/11,852, 43.8%) [33]. However, the excluded non–health-related tweets differed from those on statins. People tweeting on the non–health-related aspects of anti-diabetic drugs often referred to cost or unfair pricing, while non–health-related tweets on statins were often cultural references, jokes, financial or news reports, or web-based pharmacies.

Within our health-related tweets, it was possible to identify whether the person tweeting was discussing their own diabetes in 24.9% of the cases (415/1664), someone known to them with diabetes in 17.9% of the cases (298/1664), or if they were in a health care profession (45/1664, 2.7%). Interestingly, with those tweeting on statins [33], it was possible to identify whether the person tweeting was taking statins in 32.8% of the cases (1707/5201), someone they know taking statins in 6.6% of the cases (346/5201), or whether the person was a health care professional (325/5201, 6.2%). The much higher proportion of people discussing someone known to them with diabetes may be because of the large scale concern for people with diabetes not being able to afford their insulin.

While type 2 diabetes makes up 90% of the global cases of diabetes [1], for those tweets where we could decipher the type of diabetes more were from people with type 1 than from people with type 2 diabetes and in line with this, insulin was by far the most discussed drug (9107/9793, 92.9% of the tweets). When considering that 44.7% of the people with type 1 diabetes are younger than 40 years compared to just 4% of the people with type 2 diabetes [34] and two-thirds of Twitter users are younger than 35 years [35], a possible partial explanation is that the Twitter demographic is more aligned with the younger demographic with type 1 diabetes. Another explanation is the high proportion of people discussing the injustice of the high cost of insulin for type 1 diabetes.

The implications of high-cost insulin were far reaching. While tweets reporting bankruptcy, stealing, and homelessness associated with the cost of insulin may seem like extreme subjects to post on a public platform, a study in 2020 with individuals with type 1 diabetes in the United States corroborated these stories [36]. Approximately 39.2% of the patients struggling to afford their insulin do not tell their health care professionals [37], making Twitter a potential way of identifying patients in need. Tweets about the increasing cost of insulin reflect the general trend in the United States. The price of insulin glargine—the most commonly prescribed insulin in the United States [38]—increased by 117% over 7 years [39]. Even for those who have a Medicare insurance plan, diabetes-related out-of-pocket spending increased by 10% per
year between 2006 and 2013 [40]. This is despite the average spending for other prescription medications only increasing by 2.8% over the same period [40]. An analysis of the tweets about statins found that only 3.5% (182/5201) of the tweets mentioned cost [33] compared to 40.2% (669/1664) of the tweets in this study. This may be because the cost of a month’s supply of statins, on average, is only one-third of the price of a month’s supply of anti-diabetic drugs [41].

A relationship between cost and availability, adherence, safety and efficacy was apparent from the tweets. Twitter appeared to be an informal marketplace for trading anti-diabetic drugs, although we did not confirm actual transactions. The overall sentiment of the tweets is that the lack of affordable anti-diabetic drugs is unfair and detrimental to health, which is in agreement with the findings of Litchman et al [42], who reported that those giving away their extra anti-diabetic drugs did so out of altruism and frustration at the lack of pricing regulations rather than the need to profit. Some tweeters travelled abroad to purchase their anti-diabetic drugs; these tweeters are among the estimated 2.3 million US individuals who buy their medications abroad [43].

Although this analysis cannot quantify how many individuals do this, it does give an insight into the reasons specific to anti-diabetic drugs. Prior research has found that those without health insurance are most likely to purchase prescription medications abroad [43], and this was reflected in the tweets. Of note, Hong et al [43] inferred that those seeking health information on the internet or using web-based chat groups were twice as likely to purchase medications abroad; therefore, given that this is a Twitter analysis, there may be an overrepresentation of individuals who purchase their anti-diabetic drugs in this way. It is currently illegal to purchase insulin abroad and import it into the United States for personal use [44]; therefore, the fear of being caught may explain why there has been little mention of this method in previous studies. In July 2019, the Food and Drug Administration proposed the Safe Importation Action Plan, intending to facilitate the import of medications from Canada [45]. Despite the tweet collection covering this period, there were no tweets related to this, questioning how far this announcement spread. The tweet collection period coincided with several delays to the date the United Kingdom was due to leave the European Union. Tweets related to this highlighted the importance of protecting medication imports. The worries about imports are supported by Holt et al [46], who noted that only animal insulin is manufactured in the United Kingdom, with Novo Nordisk, Eli Lilly, and Sanofi having to import their insulins.

This study indicates the potential impact of high-cost insulin and concerns about availability, leading to rationing. This is in line with the results of a global survey of 1478 individuals with type 1 diabetes, and their care providers reported that 25.9% of the respondents from the United States had rationed their insulin at some point in the last year [47]. Rationing is deeply problematic and there was a little debate regarding insulin’s effectiveness, with powerful descriptions of how it is lifesaving. Participants with type 1 diabetes in a previous study described insulin as “life or death” for them [36], but this analysis shows that the general public also appreciates the life-saving nature of insulin. We found little evidence of the stigma associated with being on insulin among people with type 1 diabetes, which has been reported in previous studies [48]. The growing empathy for people with type 1 diabetes because of the high prices of insulin may be interconnected with a decline in the stigma.

Opinions on the efficacy of anti-diabetic drugs to treat type 2 diabetes were more varied; many tweeters expressed their desire to stop their medication, and tweets discussing other treatment options for type 2 diabetes seemed to favor dietary changes. Other studies have also indicated poor adherence in type 2 diabetes [49]. With respect to type 2 diabetes, people experience more stigma when on insulin than when on a noninsulin anti-diabetic drug [50]. A qualitative systematic review found that health care providers often doubt their patients’ ability to self-manage their diabetes, consequently preferring a paternalistic approach [51]. This is reflected in the sense of annoyance among the tweeters at not being given the option to manage type 2 diabetes by lifestyle changes alone.

There has been interest in using Twitter as a source for collecting anecdotal accounts of adverse drug reactions [13]. In our analysis of statins [33], we identified 6.8% (353/5201) of the tweets to be about adverse reactions compared to just 1.3% (21/1664) in this study. This was unexpected, given that dose-related serious adverse effects with drugs to treat diabetes are considered to be among the adverse drug effects with the highest public health impact [52], while statins have a much higher degree of safety. The cheap version ReliOn (Walmart insulin) was the only type of insulin that had its efficacy and safety questioned.

A major source of criticism of social media is the high volume of misinformation. Misinformation on social media can have detrimental effects on health behaviors, and they are difficult to correct once they gain acceptance [53]. We found little evidence of misinformation among our tweets, and in line with the literature, no misinformation was shared by health care professionals [53]. Broadly, there were 2 ways individuals used Twitter to discuss anti-diabetic drugs. The first was as a microblogging site for recording day-to-day experiences such as trying to afford their insulin, rationing, side effects, and incidences involving stigma. These tweets may provide a useful introduction into what life is like while taking anti-diabetic drugs, which could influence the support provided by health care professionals. Alternatively, Twitter was used as a tool that was intended to bring about change, with tweeters discussing complex social issues. This is pertinent to policymakers as it highlights the issues that both patients and the public consider most pressing.

**Strengths and Limitations**

The large volume of Twitter data from a mix of tweeters with and without diabetes allowed an insight into a broad range of perspectives. Manual coding was used during the tweet analysis, which is considered the gold standard method [28]. While the use of automated computer programs may be quicker and can allow large data sets to be coded, they are associated with lower accuracy [22]. These findings represent the perspectives of the Twitter-using population but not necessarily the general population [54]. As an illustration, in the United States, the average tweeter is likely to be White, young, well-educated,
and a Democrat [54]. As this study did not collect demographic data, it is hard to appreciate which population this study does reflect. Since Twitter is available worldwide, this study planned to take a global approach to anti-diabetic drug perceptions, but upon analysis, it became evident that a large burden of the tweets centered around issues in the United States. It was only after the research process began that Patel et al [55] published their analysis of 50,286 diabetes-related tweets, indicating that 43.6% of the tweets came from the United States, followed by 14.9% from the United Kingdom. Despite the large volume of tweets, we only identified issues relevant to a few countries and were unable to compare differences among countries, as we did not collect the geolocations of the Twitter users. Future work could address this. The limited non-US issues collected may, in part, be because of the search terms we used and that we only used a single social media platform. Other platforms may be explored to discover perceptions from a wider population and in other countries. Our analysis does not go beyond content analysis. We did not record any user engagement metrics or interactions. We were also unable to verify any of the claims made, and people may post things on the internet that they would not say in person. However, the fact that information shared on social media is expressed spontaneously in an open digital space with a flat role hierarchy is a major advantage for capturing perceptions that otherwise would not be reported [56]. Finally, we were unable to distinguish whether posts were referring to type 1 or type 2 diabetes in the majority of the tweets. Issues with anti-diabetic drugs are likely to be dependent on the type of diabetes. This limitation may be generalizable to other medications studied on social media, which are used for more than one indication.

**Conclusion**

The use of Twitter has provided an insight into the immediate perceptions of anti-diabetic drugs outside of a clinical setting, thereby giving a unique perspective. Not only does this study support the findings already established in the current literature, but it has also provided an appreciation of the struggles of people taking anti-diabetic drugs, particularly in light of the high cost of insulin. This study has also shown that the public is aware of these issues and are waiting for governments and health care systems to make changes.

**Acknowledgments**

This work was supported by National Institutes of Health (NIH) National Library of Medicine under grant number NIH NLM 1R01. NIH National Library of Medicine funded this research but were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. SG and KO had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Conflicts of Interest**

Sean Hennessy has received grant support and has consulted for numerous pharmaceutical companies. All other authors report no conflicts of interest.

**Multimedia Appendix 1**

Key terms used for the search.

[DOCX File, 16 KB - diabetes_v6i1e24681_app1.docx ]

**References**


19. Health language processing lab at the Institute for Biomedical Informatics. URL: https://healthlanguageprocessing.org [accessed 2020-09-16]


41. T1 International. Costs and rationing of insulin and diabetes supplies: findings from the 2018 T1 International patient survey. URL: https://pdfs.semanticscholar.org/ab82/d6c5d107dbd5c81e49d1e6477b99767ceb.pdf?_g=2.15937203.488990161.1549582198.1434567109.15494038301] [accessed 2020-07-13]
©Su Golder, Millie Bach, Karen O'Connor, Robert Gross, Sean Hennessy, Graciela Gonzalez Hernandez. Originally published in JMIR Diabetes (http://diabetes.jmir.org/), 26.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Review

Application of the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies in Assessing Mobile-Delivered Technologies for the Self-Management of Type 2 Diabetes Mellitus: Scoping Review

Jessica R Forsyth¹, BA; Hannah Chase¹, MA VetMB; Nia W Roberts², MSc; Laura C Armitage¹, MB BCh, MRCGP; Andrew J Farmer³, DM, FRCGP

¹Medical Sciences Division, University of Oxford, Oxford, United Kingdom
²Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom
³Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Laura C Armitage, MB BCh, MRCGP
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building
Radcliffe Observatory Quarter, Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865 617942
Email: laura.armitage@phc.ox.ac.uk

Abstract

Background: There is a growing role of digital health technologies (DHTs) in the management of chronic health conditions, specifically type 2 diabetes. It is increasingly important that health technologies meet the evidence standards for health care settings. In 2019, the National Institute for Health and Care Excellence (NICE) published the NICE Evidence Standards Framework for DHTs. This provides guidance for evaluating the effectiveness and economic value of DHTs in health care settings in the United Kingdom.

Objective: The aim of this study is to assess whether scientific articles on DHTs for the self-management of type 2 diabetes mellitus report the evidence suggested for implementation in clinical practice, as described in the NICE Evidence Standards Framework for DHTs.

Methods: We performed a scoping review of published articles and searched 5 databases to identify systematic reviews and primary studies of mobile device–delivered DHTs that provide self-management support for adults with type 2 diabetes mellitus. The evidence reported within articles was assessed against standards described in the NICE framework.

Results: The database search yielded 715 systematic reviews, of which, 45 were relevant and together included 59 eligible primary studies. Within these, there were 39 unique technologies. Using the NICE framework, 13 technologies met best practice standards, 3 met minimum standards only, and 23 technologies did not meet minimum standards.

Conclusions: On the assessment of peer-reviewed publications, over half of the identified DHTs did not appear to meet the minimum evidence standards recommended by the NICE framework. The most common reasons for studies of DHTs not meeting these evidence standards included the absence of a comparator group, no previous justification of sample size, no measurable improvement in condition-related outcomes, and a lack of statistical data analysis. This report provides information that will enable researchers and digital health developers to address these limitations when designing, delivering, and reporting digital health technology research in the future.

(JMIR Diabetes 2021;6(1):e23687) doi:10.2196/23687

KEYWORDS

type 2 diabetes; health technology; self-management; mobile health; mobile applications; guidelines
Introduction

Background

Digital technologies are now integral to the delivery of health care and feature in policies for the future of national [1] and global [2] health care systems. The World Health Organization (WHO) defines a health technology as “the application of organized knowledge and skills in the form of devices, medicines, vaccines, procedures and systems, developed to solve a health problem and improve quality of lives” [3]. Typically, digital health technologies (DHTs) include apps, software, and web-based platforms intended to benefit people or the wider health care system [4]. DHTs are increasingly supporting or being used as an adjunct to face-to-face clinical care by facilitating remote health care.

Many DHTs are intended to support chronic disease management, where self-management and preventative medicine are key components of effective care. Approximately 500 million people use mobile device apps to manage their health [5], and diabetes is the condition most commonly targeted by commercial apps [6]. With an increasing global prevalence of type 2 diabetes, mobile device apps offer a potential means of supporting diabetes care, particularly in the context of increasing demands against limited resources. It is imperative that the quality, safety, and effectiveness of such mobile device apps are assessed before deployment in clinical practice. In 2019, the WHO cautioned that amid increasing interest, digital health has been characterized by interventions being implemented without careful examination of the evidence base on their benefit and harms [7]. In the same year, the National Institute for Health and Care Excellence (NICE) published the Evidence Standards Framework for DHTs to guide clinicians, researchers, and policy makers in assessing whether the published literature evaluating these technologies provides the required level of evidence for their intervention to be considered for use in the UK health care setting [4].

There are several existing guidelines on evaluating the use of DHTs, including guidelines by policy makers such as the WHO, the United States’ Federal Drug Administration, and National Health Service England [8-11] as well as frameworks developed by independent research groups [12,13]. However, the NICE framework is unique in explicitly suggesting a quality standard in relation to a technology’s functionality. Although the NICE framework was developed for DHTs used in a UK health care setting, the framework has the advantage of being research oriented rather than reliant on nation-specific commercial standards. This provides an opportunity for applying the framework to broader settings. First, the research-based focus may allow the framework to be used to evaluate the effectiveness of both consumer-driven and clinician-prescribed DHTs. Second, the framework may also be adapted to other health care systems by adjusting the requirement for development and testing in the United Kingdom to that of the DHT’s host country. Therefore, the NICE Evidence Framework may be used to guide assessment of and make comparisons between scientific literature regarding a variety of DHTs developed and applied internationally.

The NICE framework classifies apps by function and stratifies them into tiers (tiers 1, 2, 3a, or 3b). The tier framework corresponds with the evidence level required to support use of the technology; requirements are cumulative, becoming increasingly rigorous from tier 1 to 3 and divided into best practice and minimum standards. Stakeholders are encouraged to assess the evidence against these standards, which include, for example, whether the study measures important outcomes for users, whether the intervention works independently of health care professionals’ input, and the extent to which the intervention guides diagnosis, management, and treatment of a disease.

To date, there has been no review exploring whether peer-reviewed scientific literature regarding DHTs meets these evidence requirements. We investigated this in the context of DHTs designed to support the self-management of type 2 diabetes, as it is the most common chronic condition targeted by self-management DHTs [6].

Objectives

The objectives of this review are (1) to systematically identify peer-reviewed publications on mobile device DHTs intended to support or encourage the self-management of type 2 diabetes mellitus (T2DM), (2) to use the NICE Evidence Standards Framework to allocate each DHT to the appropriate intervention tier based on their described technology and function, and (3) to examine the extent to which the evidence reported for the identified DHTs meets the NICE framework level of evidence required according to its tier.

Methods

Review Design

We performed a scoping review [14] to understand the literature to date and explore the application of research methodology in relation to the NICE evidence standards. The review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [15].

Data Sources

A total of 5 databases (MEDLINE, Embase, PsycINFO, CINAHL, and Cochrane Database of Systematic Reviews) were searched for systematic reviews published between January 2000 and August 2019 that evaluated mobile device DHT interventions for T2DM. Our database choice and search strategy were developed through consultation with a medical information specialist to identify the most relevant sources for peer-reviewed medical and clinical research studies. An example search strategy is provided in Multimedia Appendix 1.

Screening for Systematic Reviews

Two reviewers (JF and LA) independently screened all citations for systematic reviews by title and abstract and excluded those that clearly did not meet the eligibility criteria. Decisions were then unblinded, and any conflicting decisions were arbitrated by a third reviewer (AF). Full-text articles for all included citations were then screened against the inclusion criteria by 2 reviewers (JF and LA).
Reviews were eligible if they included primary studies evaluating mobile apps designed to support adults with the self-management of diabetes mellitus. Reviews were excluded if they included studies in which the study population included people with type 1 diabetes, an undifferentiated mix of people with type 1 diabetes or type 2 diabetes, gestational diabetes, childhood diabetes or prediabetes, or focused on diagnosing diabetes (due to our focus on assessing DHTs designed to support self-management). Reviews that focused exclusively on telemedicine or telehealth interventions were also excluded, owing to our focus on technologies that support self-management and therefore require some degree of functionality independent of a clinician.

Screening for Primary Studies and Technologies

Relevant primary studies were then identified from eligible systematic reviews. The eligible reviews were equally divided between the 4 reviewers (JF, LA, HC, and AF) who then screened the title and abstract of each primary study included in each review. When a primary study was excluded, the study was double screened by a second reviewer, and in the instance of any conflict, a third reviewer arbitrated (LA or AF). Primary studies included at this stage were then divided between the 4 reviewers who reviewed the full text of each study for eligibility. Furthermore, when a study was excluded, the study was double screened by a second reviewer, and any conflict was arbitrated by a third reviewer (LA or AF).

Primary studies were eligible for inclusion if they met the following inclusion criteria:

1. Population: adults with a diagnosis of T2DM.
2. Intervention: a mobile device–delivered DHT designed to support the self-management of T2DM, which provides support independent of a clinician.

Data Extraction

Data were extracted from the included primary studies by 4 reviewers (JF, LA, HC, and AF). We designed a custom data extraction form using the evidence for effectiveness tables from the NICE framework [4] and additional guidance in the framework; an explanation of this approach can be found in Multimedia Appendix 2.

We extracted the following items from primary studies: (1) DHT investigated, (2) year of study, (3) study nation, (4) study design, (5) study setting, (6) outcomes of interest, (7) study duration and follow-up period, (8) sample size, (9) recruitment setting, (10) comparator group, (11) improvement in outcome with intervention, (12) justification of sample size, (13) statistical methods, and (14) follow-up rate. For tier 3a studies, we also extracted the following item: (15) description of and reference to a behavior change technique. Where more than one article that investigated the same DHT intervention was identified, data were extracted separately for each article.

Assigning Technologies and Intervention Tier

Descriptions of each technology were extracted from the primary studies, and we assigned each app a tier according to the NICE framework, as described in Multimedia Appendix 2. Where an app had more than one function, the function with the highest applicable tier was considered when assigning an overall tier. Tier 3b was considered as a higher tier to 3a owing to its more rigorous evidence requirements, as detailed in Multimedia Appendix 2.

Assessment of Evidence According to Tier

We used the NICE framework to evaluate each DHT against evidence levels, referring to evidence in the primary studies for each DHT, as described in Multimedia Appendix 2. We assessed each technology against its highest relevant tier to determine whether the DHT met the framework’s minimum and best practice evidence requirements. Where a technology was reported in more than one primary study, we analyzed each primary study separately against the framework and selected the strongest supporting evidence for the technology reported across the primary studies.

We also compared the NICE evidence standards outcome for a DHT against the income status of the study nation (as defined by the World Bank [16]). This was done to explore whether the NICE framework could be applied to DHTs designed for a different health care structure and system outside of the United Kingdom; a need for more empirical approaches to assess DHTs in low- and middle-income countries has been highlighted in recent literature [17,18].

Tier 3a guidance requires evidence of a referenced behavioral change technique (BCT) in the development or use of a technology that encourages behavioral change. For the purposes of this review and evidence assessment, we took a pragmatic decision to exclude this requirement in our overall decision on whether a tier 3a technology met the evidence requirements, accounting for the fact that our search methods may not have identified all relevant development studies reporting on a technology’s design.

In addition, the framework defines data quality as the presence of “statistical considerations such as sample size and statistical testing.” A pragmatic decision was made that statistical testing of some degree was needed as the minimum evidence requirement for all studies. However, the framework accommodates observational and quasi-experimental study designs, where it is impractical to statistically justify the sample size. Therefore, when making an assessment of evidence for studies of these designs, a statistical justification of sample size was not needed to meet minimum standards (but was required for experimental studies or randomized controlled trials [RCTs]).

Results

Screening for Systematic Reviews

The initial database search returned 715 citations. After removal of duplicates, 709 citations were screened by title and abstract. We identified 68 relevant systematic reviews for which we screened the full-text articles. Of these, 45 reviews were included (Figure 1).
Screening for Primary Studies and Technologies

From these 45 reviews, we identified 145 relevant primary studies and screened their full-text articles. Of these, 61 primary studies met the inclusion criteria described above. We subsequently excluded 2 studies because there was insufficient information describing their technology to allocate a tier. The remaining 59 studies described 39 unique technologies and were included for data extraction (Figure 1).

The characteristics of the 59 included studies are presented in Multimedia Appendix 3 [19-77]. The publication year of the included studies ranged from 2007 to 2017. Of the included 59 studies, 36 (61%) were RCTs (of which 7 were identified as feasibility or pilot studies) and 23 (39%) were observational cohort studies (of which 19 were identified as feasibility or pilot studies). Qualitative data were reported alongside 6 RCTs and 13 observational cohort studies. The study nation varied, with 23 studies conducted in the United States, 6 in Norway, 4 in Korea, 3 studies each in Canada, the United Kingdom, and Saudi Arabia, 2 studies each in the Netherlands, Japan, Iran, and India, and 1 study each in Singapore, Mexico, Finland, Iraq, Bangladesh, the Democratic Republic of Congo, and China. Of the 39 technologies included for data analysis, 17 (44%) were mobile apps, 2 (5%) were personal digital assistant apps, and 20 (51%) were automated SMS.

Assigning Technologies to an Intervention Tier

All DHTs identified and included in this review were classified as tier 3 technologies. Descriptions of the technologies and their assigned subtiers are presented in Table 1 for tier 3a and Table 2 for tier 3b.

Of the 39 technologies, 23 (59%) were assigned to tier 3a. Tier 3a describes DHTs used for preventing and managing diseases and is divided into preventative behavior change and self-manage. Of these 23 technologies, 6 were apps and 17 were SMS based. Of the tier 3a technologies, 12 were classified as preventative behavior change only, 3 were classified as self-manage only, and 8 had both 3a preventative behavior change and self-manage characteristics.

We assigned 16 (41%) of the 39 technologies to tier 3b. Tier 3b describes technologies used as tools for treatment, diagnosis, and management decisions and is divided into treat, active monitoring, calculate, and diagnose. Of these 16 technologies, 13 were apps and 3 were SMS based. Of the tier 3b technologies, 7 were active monitoring only, 3 were treat and active monitoring, 1 was treat and calculate, 1 was active monitoring and calculate, and 4 had all 3 of the 3b treat, active monitoring, and calculate characteristics.
<table>
<thead>
<tr>
<th>Digital health technology and description</th>
<th>Self-manage</th>
<th>PBC&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tier 3a app technologies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes Pilot [19-22]</td>
<td>✓</td>
<td>N/A&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Few Touch App (FTA) [23-28]</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Unnamed (Sevick) [29]</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Monica [30]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>iDecide [31]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Diabetes 101 [32]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><strong>Tier 3a SMS technologies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NICHE system [33]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Shetty) [34]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Diabetech [35]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Goodarzi) [36]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Real-Time Medication Monitoring [37,38]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Care4Life [39,40]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SMS-DMCare [41]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>MEssaging for Diabetes (MED) [42]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>TExT-MED [43,44]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Haddad) [45]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Argay) [46]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Bin Abbas) [47]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Islam) [48]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Text to Move [77]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Peimani) [49]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Unnamed (Fang) [50]</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Dulcedigital [51]</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

<sup>a</sup>PBC: preventative behavior change.

<sup>b</sup>PDA: personal digital assistant.

<sup>c</sup>HCP: health care professional.

<sup>d</sup>Digital health technology falls within the subtier.
eN/A: not applicable.
fHbA1c: glycated hemoglobin.
gT2DM: type 2 diabetes mellitus.
hBG: blood glucose.
Table 2. Tier 3b digital health technologies: descriptions and subtier allocation (N=16).

<table>
<thead>
<tr>
<th>Digital health technology and description</th>
<th>Treat</th>
<th>Active monitoring</th>
<th>Calculate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tier 3b app technologies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP&lt;sup&gt;a&lt;/sup&gt; telemangement [52]</td>
<td></td>
<td>✓✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient BP automatically uploaded. HCP&lt;sup&gt;b&lt;/sup&gt; accesses all data. Alert to patient and HCP if critical. Automatic BP reminders to patient</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>✓&lt;sup&gt;d&lt;/sup&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td>WellDoc [53-58]</td>
<td></td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient BG automatically uploaded, medication dose and diet self-inputed: automated personalized feedback on medication dose and behavior. HCP accesses all data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>t+ Diabetes [59-61]</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient BG automatically uploaded and insulin dose self-inputed: displayed graphically, decision aids for self-titration. HCP accesses all data and messages through the app</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobil Diab [62]</td>
<td></td>
<td>✓✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient BG automatically uploaded, displayed graphically. HCP accesses all data and sends feedback through the app</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Health Coach App [63,64]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient BG automatically uploaded, displayed graphically. HCP accesses all data, individualized feedback, and two-way communication through the app</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Dialbetics app [65,66]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient self-inputs BG data: behavioral feedback and alerts if abnormal. HCP accesses all data; abnormal readings flagged. Features: later version includes dietary feedback</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>SANAD [67]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: BG&lt;sup&gt;e&lt;/sup&gt; automatically uploaded. Features: social networking module and CBT&lt;sup&gt;f&lt;/sup&gt; module. HCP accesses all data; sends feedback through app</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>SAED system [68]</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: BG automatically uploaded. Features: weekly educational message. HCP accesses all data; two-way communication through the app</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Diabetes Pal [69]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient self-inputs BG: app suggests insulin dose (within the preset range). Features: educational information. Research staff access all data; flag to HCP</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
</tr>
<tr>
<td>CollaboRhythm [70]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient self-inputs medication and BG displayed graphically. HCP accesses all data and suggests insulin correction; two-way communication through the app</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PSDCS [71]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: BG automatically uploaded, diet and exercise self-inputed—feedback and suggested insulin changes based on algorithm. Features: automated daily recommendations for calorie intake and exercise</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Brew app [72]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile app: patient self-inputs health data. Features: daily SMS reminders, educational information. HCP accesses summary of data and sends alerts for BG or missed appointments</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Gather Health [73]</td>
<td></td>
<td>✓✓</td>
<td></td>
</tr>
<tr>
<td>Mobile app: patient self-inputs BG: displayed graphically. Features: daily reminders and self-care advice. HPC accesses all data; two-way communication through the app</td>
<td>N/A</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Tier 3b SMS technologies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCDC system [74]</td>
<td></td>
<td>✓✓</td>
<td>N/A</td>
</tr>
<tr>
<td>SMS: patient BG automatically sent to server, automated summary SMS with behavioral suggestions. Patient sends BP and exercise via SMS. Informational SMS trice daily. HCP accesses all data</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Unnamed SMS (Kim) [75]</td>
<td></td>
<td>✓✓</td>
<td></td>
</tr>
<tr>
<td>SMS: patient BG automatically sent to server, automated SMS suggestions to adjust insulin based on an algorithm. If hypoglycemic, emergency SMS sent to patient and caregiver</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CDSS u-health care [76]</td>
<td></td>
<td>✓✓</td>
<td></td>
</tr>
<tr>
<td>SMS: Patients BG automatically uploaded to server, automated daily SMS summaries, suggestions to adjust insulin based on algorithm, weekly and monthly summaries</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<sup>a</sup>BP: blood pressure.

<sup>b</sup>HCP: health care professional.

<sup>c</sup>N/A: not applicable.

<sup>d</sup>Digital health technology falls within the subtier.
Assessment of Evidence According to Tier

The assessment of evidence level according to the assigned tier is presented in Table S1 [22,28-36,38,41-43,45-51,77] in Multimedia Appendix 4 for tier 3a technologies and in Table S2 [52,54,61,62,64,65,67-76,78] in Multimedia Appendix 4 for tier 3b technologies. Across all 39 technologies, 11 demonstrated best practice standards for the evidence level assigned, 3 technologies demonstrated minimum standards, and 25 did not report methods or findings that met minimum standards.

Tier 3a Technologies

Of the 23 tier 3a technologies, 7 met the best practice standards, 3 met the minimum evidence standards, and 13 did not report methods or findings reaching minimum standards. Of the 13 technologies that did not provide evidence for minimum standards, there were several common reasons for falling short of the minimum standard. First, 7 technologies did not provide statistical justification of sample size where the study design was appropriate, with this being the only reason for not meeting minimum standards in all 7 technologies. Second, 6 technologies did not provide comparative data, with this being the only reason for not meeting the minimum standards in the 2 technologies. Finally, 3 technologies did not conduct any statistical testing on the data set.

For the 3 tier 3a technologies that met the minimum evidence standards, there were 2 common reasons why these technologies did not meet the best practice standards. First, 2 technologies showed no improvement in condition-relevant outcomes, with this being the only reason for both technologies not meeting the best practice. Second, 1 technology’s comparator group did not represent usual care, with this being the only reason for not meeting the best practice.

Tier 3b Technologies

Of the 16 tier 3b technologies, 4 met best practice standards, none met only minimum evidence standards, and 12 did not report methods or findings reaching minimum standards. Of the 12 technologies that did not provide evidence for minimum standards, there were several common reasons for falling short of the minimum standard. First, 3 technologies used a single-arm cohort study design that lacked a comparator group and failed to meet the requirement of design being quasi-experimental or higher, with inappropriate study design being the only reason for not meeting minimum standards in all 3 technologies. Second, 7 technologies had no statistical justification of sample size where the study design was appropriate, with this being the only reason for 5 of these technologies. Third, there were 2 technologies that did not conduct any statistical testing on the data set. Finally, 2 technologies had a follow-up period of less than 3 months, which is the accepted minimum clinically relevant follow-up period for type 2 diabetes.

Evidence Standard by Host Country

Table 3 shows the DHTs arranged according to the income status (as defined by the World Bank [16]) of the study nation and the outcome of the DHT’s NICE evidence assessment. There were considerably more DHTs from high-income economies (n=30) than upper middle-income (n=5), lower middle-income (n=3), or low-income (n=1) economies. In addition, there was no evidence of studies from high-income nations being more or less successful in meeting NICE evidence standards than lower-income nations: only 9 out of 30 DHTs investigated in high-income economies met either minimum or best practice standards, compared with 3 out of 5 DHTs investigated in upper middle-income economies, 2 out of 3 DHTs investigated in low- and middle-income economies, and 0 out of 1 DHTs investigated in low-income economies.
Table 3. Digital health technologies arranged by World Bank income status of host country and the digital health technology evidence outcome (N=39).

<table>
<thead>
<tr>
<th>Country</th>
<th>DHT*</th>
<th>NICE* evidence level met</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low-income economies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>Mobil Diab</td>
<td>No</td>
</tr>
<tr>
<td><strong>Lower middle-income economies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>Unnamed (Islam)</td>
<td>Best practice</td>
</tr>
<tr>
<td>India</td>
<td>Unnamed (Shetty)</td>
<td>No</td>
</tr>
<tr>
<td>India</td>
<td>Gather Health</td>
<td>Best practice</td>
</tr>
<tr>
<td><strong>Upper middle-income economies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Unnamed (Fang)</td>
<td>Minimum</td>
</tr>
<tr>
<td>Iran</td>
<td>Unnamed (Haddad)</td>
<td>No</td>
</tr>
<tr>
<td>Iran</td>
<td>Unnamed (Goodarzi)</td>
<td>Best practice</td>
</tr>
<tr>
<td>Iraq</td>
<td>Unnamed (Peimani)</td>
<td>Best practice</td>
</tr>
<tr>
<td>Mexico</td>
<td>Brew app</td>
<td>No</td>
</tr>
<tr>
<td><strong>High-income economies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>BP telemanagement</td>
<td>No</td>
</tr>
<tr>
<td>Canada</td>
<td>Health Coach App</td>
<td>No</td>
</tr>
<tr>
<td>Finland</td>
<td>Monica</td>
<td>No</td>
</tr>
<tr>
<td>Hungary</td>
<td>Unnamed (Argay)</td>
<td>No</td>
</tr>
<tr>
<td>Japan</td>
<td>Diabetics app</td>
<td>Best practice</td>
</tr>
<tr>
<td>Korea</td>
<td>CDSS-based u-health care</td>
<td>No</td>
</tr>
<tr>
<td>Korea</td>
<td>PSDCS</td>
<td>No</td>
</tr>
<tr>
<td>Korea</td>
<td>UCDC system</td>
<td>No</td>
</tr>
<tr>
<td>Korea</td>
<td>Unnamed (Kim)</td>
<td>Best practice</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Real-Time Medication Monitoring</td>
<td>No</td>
</tr>
<tr>
<td>Norway</td>
<td>Few Touch Application</td>
<td>Minimum</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>SANAD</td>
<td>No</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>SAED</td>
<td>No</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>Unnamed (Bin Abbas)</td>
<td>No</td>
</tr>
<tr>
<td>Singapore</td>
<td>Diabetes Pal</td>
<td>No</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>t+Diabetes</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Care4life</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>CollaboRhythm</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Diabetech</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Dulcedigital</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Diabetes 101</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>MED</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>NICHE system</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>SMS-DMCare</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Unnamed (Sevick)</td>
<td>Minimum</td>
</tr>
<tr>
<td>United States</td>
<td>Diabetes Pilot</td>
<td>Best practice</td>
</tr>
<tr>
<td>United States</td>
<td>iDecide</td>
<td>Best practice</td>
</tr>
<tr>
<td>United States</td>
<td>TEXT-MED</td>
<td>Best practice</td>
</tr>
<tr>
<td>United States</td>
<td>Text to Move</td>
<td>Best practice</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

We aimed to evaluate whether peer-reviewed literature investigating the use of mobile device DHTs for the self-management of T2DM met the required evidence level set out in the NICE Evidence Standards Framework for DHTs. The framework aims to ensure that new technologies introduced to clinical health care settings are effective and offer economic value. We identified 39 mobile device DHTs designed to support self-management of T2DM in the scientific literature; these were a mix of app-based and SMS-based technologies. We found that all technologies fell into tier 3a or tier 3b (the highest tiers) of the NICE framework, with tier 3 interventions targeting disease management and requiring the most rigorous evidence. When assessing a technology using the NICE Evidence Standards Framework, we assessed all primary studies supporting a DHT individually against the framework and selected the strongest supporting evidence for the technology reported across the primary studies.

For more than half of the technologies identified, the underpinning literature did not meet the evidence standards to demonstrate effectiveness, as recommended by the NICE framework for the technology’s tier. Of the 39 technologies identified, only 16 met minimum or best evidence standards, with 23 not meeting the minimum requirements. The most common reasons for not meeting the NICE standards included a lack of an appropriate comparator group that reflected usual care, no statistical justification of sample size, a lack of measurable improvement in condition-related outcomes, and no statistical data analysis. Given the high proportion of RCTs among the identified studies (36/59, 61%), it was surprising that such a large number did not meet the minimum evidence standards due to these reasons. We found that the evidence framework could easily be applied to a variety of study nations and that studies from a range of economic settings were able to meet evidence standards for the DHT. From the results of this study, we suggest that the application of DHT evidence standards are globally relevant.

Using the NICE Evidence Standards Framework to Evaluate Evidence

We encountered several challenges in interpreting and using the NICE framework. First, we found that for diabetes, there was ambiguity in distinguishing technology for healthy living and technology for disease management. The same technology that targeted diet and exercise could be considered tier 2 for people without diabetes as a healthy living app but tier 3 for those with T2DM as a disease management app. There are several terms used in the NICE framework that can be ambiguous in their application and may require greater clarity, including the phrases high quality data and clinically relevant follow-up period. The framework does not include guidance as to how either of these points should be assessed.

As the NICE Evidence Framework was designed in the United Kingdom, the standards reference the UK health care setting when assessing the development and effectiveness of a technology. We found that adaptation of the NICE framework to assess a DHT in its host country, rather than specifically in the United Kingdom, allowed the analysis and comparison of DHTs in an international context. We also noted that the UK-specific requirement may restrict UK policy makers, commissioners, and clinicians from adopting and implementing DHTs that have been rigorously evaluated in another health care setting and do not require substantial adaptation. This could be considered overly restrictive for DHTs that target self-management and may not need integration with a health care system.

Finally, we observed a potential mismatch between the level of risk associated with an intervention and the level of evidence required according to the intervention’s associated tier. For example, Real-Time Medication Monitoring [37,38], which would be categorized under tier 3a (preventative behavior change due to explicit suggestions by the DHT to the patient for actions or behavior change) might be considered a low-risk technology, involving automatic SMS reminders to take medication when a patient’s pill box remains unopened. However, Health Coach App [63,64], also classified under tier 3a (self-management for symptoms, health or disease related data, or medication tracking over time) might be considered as having higher risk, tracking multiple health behaviors, holding sensitive data, and facilitating two-way messaging. Despite this difference in the level of risk, both technologies fall under the same tier and require the same standard of supporting evidence. The evidence framework also stipulates that any technology where there is automatic transfer of data (regardless of type) to a health care professional should be categorized as tier 3b rather than tier 3a under active monitoring, requiring more rigorous evidence for clinical input without any apparent additional risk. Therefore, tier levels may need to be adjusted to reflect clinical risk rather than function alone.

Strengths and Limitations

Although this is a scoping review, we took a systematic approach to identify peer-reviewed articles, adding rigor to our methods. We included reviews of all study design types, including experimental, observational, and qualitative study designs. However, while we identified several experimental and observational studies, this approach may not have captured all developmental studies and recently published studies that are less likely to be included in systematic reviews. However, we would have expected developmental studies to be cited in subsequent experimental and observational clinical studies, and we hand-searched full-text articles for such studies. We adapted
our evidence assessments where appropriate (eg, excluding requirements for BCT evidence in tier 3a).

We identified technologies that have been investigated and published in the scientific literature and did not review app catalogs or commercial publications for relevant technologies. We feel this approach was appropriate, as we did not have the resources to obtain and evaluate these sources and assess the extent to which they meet evidence standards, as described in the NICE framework. In addition, although the NICE framework was developed for DHTs used in a clinical setting, we did not differentiate between commercial and commissioned DHTs in this study. However, we encountered no challenges in applying the tier 3 evidence requirements to technologies scientifically evaluated either by clinical or commercial teams; indeed, the evidence framework could be used to design studies to evaluate the use of commercial apps within a clinical setting. Although we assessed the income status of the study nation to explore the applicability of the framework in a variety of health care settings, this did not take into account the scenario where a technology was developed in a high-income country but delivered in a low-income population [31,42-44,51,63,64]. Although beyond the scope of this review, future work could explore the effect of sociodemographic factors of the target population (such as economic status, access to health care, and technology literacy) in using the framework to evaluate the effectiveness of DHTs.

Due to potential ambiguity and subjectivity applying the NICE framework, we acknowledge that our interpretation will have affected decisions around classification and evidence evaluation and consequently the number of DHTs meeting evidence standards. We have highlighted that greater clarity of key terms in the framework would be valuable. We also acknowledge that the scope of our analysis was limited to the evidence requirements in the NICE framework, but other considerations for study quality (ie, prospective registration, retention rate) and intervention effect (ie, technology literacy, impact on behavior) are interesting and relevant in evaluating the effectiveness of DHTs.

We identified several evidence-level criteria as described by NICE that studies of DHTs commonly failed to meet. This offers a useful resource for digital health researchers and developers who may use this information in designing and reporting DHT research in the future. This might aid in the translation of research into clinical care by ensuring that the required information is measured and reported. This in turn will enable commissioners, policy makers, and clinicians to readily assess whether a technology is suitable for implementation in the UK health care setting.

Comparison With Previous Work
Previous studies have identified a lack of evidence of an effect in apps for diabetes. Recently, Veazie et al [79] identified 15 studies evaluating 11 apps for the self-management of diabetes and found that only 5 technologies were supported by evidence showing significant clinical improvement with use. Our study supported this finding as well as identifying many more apps and several other aspects of evidence that could be improved. In addition, a previous study highlighted challenges in applying the NICE Evidence Framework tiers in classifying DHTs. Nwe et al [80] used the NICE framework to classify 76 apps from the National Health Service (NHS) app library into their relevant technology tier and assessed the classification agreement between 2 mobile health (mHealth) researchers. They found a disagreement on the classified tier in 45% (34/76) of technologies [80]. Our study complements the author’s recommendation that greater clarity in the framework may be needed to improve the consistency of its application. To our knowledge, this is the first study to assess the evidence supporting DHTs against the NICE Evidence Framework. Previous reviews evaluating DHTs in other clinical settings, such as technologies for stroke rehabilitation and virtual reality tools in pediatric care, have highlighted the need for a set of recognized standards in the field with specific mention to the NICE framework [81,82]. Therefore, it would be of interest to assess and compare the application of the NICE framework with DHTs in other health care settings in addition to chronic disease management. Given that the NICE framework is relatively new, it would be valuable to conduct similar reviews in the future to assess the potential impact of the framework on rigor and quality of studies over time.

Conclusions
This review evaluated a defined group of mobile-delivered DHTs designed for use by people with T2DM, using the NICE Evidence Standards Framework for DHTs. Over half of the identified DHTs did not meet the minimum evidence standards required for their intervention tier, as defined by the NICE Evidence Standards Framework. This may pose a major barrier to the translation of mHealth interventions into the UK health care setting. However, we have highlighted the most common areas in which DHT evaluations do not meet the standards set out by NICE, and this provides an opportunity for researchers and DHT developers to address these points when designing and reporting DHTs in the future. In addition, we identified the potential scope for development of the NICE framework so that the evidence tiers correlate more closely with the associated risk of an intervention. Above all, commissioners, clinicians, and patients need to have confidence in the safety of DHTs for these to be implemented into everyday chronic disease management, and increased risk should be underpinned by the most rigorous scientific research.

Acknowledgments
This review did not receive any funding. This research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS,
the NIHR or the Department of Health. AF is an NIHR Senior Investigator, and both AF and LA receive support from the NIHR Oxford Biomedical Research Centre.

**Conflicts of Interest**

AF is Program Director of the NIHR Health Technology Assessment Programme.

--

**Multimedia Appendix 1**

Example of full search strategy for the Medline database.

[PDF File (Adobe PDF File), 734 KB - diabetes_v6i1e23687_app1.pdf]

**Multimedia Appendix 2**

An explanation of the classification strategy for digital health technologies using the technology tier and evidence level in the National Institute of Health and Care Excellence Framework.

[PDF File (Adobe PDF File), 724 KB - diabetes_v6i1e23687_app2.pdf]

**Multimedia Appendix 3**

Characteristics of primary studies included for data extraction.

[PDF File (Adobe PDF File), 589 KB - diabetes_v6i1e23687_app3.pdf]

**Multimedia Appendix 4**

Overall technology assessments against the National Institute for Health and Care Excellence Evidence Framework.

[PDF File (Adobe PDF File), 608 KB - diabetes_v6i1e23687_app4.pdf]

**References**


10. Multiple function device products: policy considerations guidance for industry food drug administration staff internet. FDA. 2018. URL: https://www.fda.gov/media/112671/download [accessed 2021-01-16]

11. Policy for device software functions mobile medical applications guidance for industry food drug administration staff preface public comment internet. FDA. 2019. URL: https://www.fda.gov/media/80958/download [accessed 2021-01-16]


http://diabetes.jmir.org/2021/1/e23687/


Abbreviations

BCT: behavior change technique
DHT: digital health technology
mHealth: mobile health
NICE: National Institute of Care Excellence
NIHR: National Institute for Health Research
NHS: National Health Service
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus
WHO: World Health Organization
Diabetes Distress and Glycemic Control in Type 2 Diabetes: Mediator and Moderator Analysis of a Peer Support Intervention

Kara Mizokami-Stout, MSc, MD; Hwajung Choi, PhD; Caroline R Richardson, MD; Gretchen Piatt, MPH, PhD; Michele Heisler, MPA, MD

1National Clinician Scholars Program, Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, United States
2Division of Metabolism, Endocrinology and Diabetes, University of Michigan, Ann Arbor, MI, United States
3Ann Arbor Veteran Affairs Hospital, Ann Arbor, MI, United States
4Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States
5Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, United States
6Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States
7Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI, United States
8Department of Health Behavior and Health Education, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Kara Mizokami-Stout, MSc, MD
Division of Metabolism, Endocrinology and Diabetes
University of Michigan
1000 Wall Street
5100 Brehm Tower
Ann Arbor, MI,
United States
Phone: 1 734 232 1269
Email: kmizokam@med.umich.edu

Abstract

Background: High levels of psychosocial distress are correlated with worse glycemic control as measured by glycosylated hemoglobin levels (HbA1c). Some interventions specifically targeting diabetes distress have been shown to lead to lower HbA1c values, but the underlying mechanisms mediating this improvement are unknown. In addition, while type 2 diabetes mellitus (T2D) disproportionately affects low-income racial and ethnic minority populations, it is unclear whether interventions targeting distress are differentially effective depending on participants’ baseline characteristics.

Objective: Our objective was to evaluate the mediators and moderators that would inform interventions for improvements in both glycemic control and diabetes distress.

Methods: Our target population included 290 Veterans Affairs patients with T2D enrolled in a comparative effectiveness trial of peer support alone versus technology-enhanced peer support with primary and secondary outcomes including HbA1c and diabetes distress at 6 months. Participants in both arms had significant improvements in both HbA1c and diabetes distress at 6 months, so the arms were pooled for all analyses. Goal setting, perceived competence, intrinsic motivation, and decisional conflict were evaluated as possible mediators of improvements in both diabetes distress and HbA1c. Baseline patient characteristics evaluated as potential moderators included age, race, highest level of education attained, employment status, income, health literacy, duration of diabetes, insulin use, baseline HbA1c, diabetes-specific social support, and depression.

Results: Among the primarily African American male veterans with T2D, the median age was 63 (SD 10.2) years with a baseline mean HbA1c of 9.1% (SD 1.7%). Improvements in diabetes distress were correlated with improvements in HbA1c in both bivariate and multivariable models adjusted for age, race, health literacy, duration of diabetes, and baseline HbA1c. Improved goal setting and perceived competence were found to mediate both the improvements in diabetes distress and in HbA1c. Baseline patient characteristics evaluated as potential moderators included age, race, highest level of education attained, employment status, income, health literacy, duration of diabetes, insulin use, baseline HbA1c, diabetes-specific social support, and depression.

Conclusions: Prior studies have demonstrated that some but not all interventions that improve diabetes distress can lead to improved glycemic control. This study found that both improved goal setting and perceived competence over the course of the
Introduction

Diabetes distress, or the negative emotional and behavioral responses that can occur as a result of having a demanding chronic illness like diabetes, is an increasingly recognized psychosocial factor influencing diabetes self-management [1]. The prevalence of at least moderate levels of diabetes distress is up to 45% in adults with type 2 diabetes (T2D) [2], and high levels of diabetes distress lead to poor medication adherence, higher glycosylated hemoglobin $A_1c$ ($HbA_1c$) values, and, ultimately, poor quality of life [2-4].

While the link between high levels of diabetes distress and higher $HbA_1c$ has been well established [1], a number of evaluated interventions specifically targeting diabetes distress lead to improvements in glycemic control [5]. Examples of such interventions include educational, psychosocial, or psychological programs (including cognitive behavioral therapy, motivational interviewing, and mindfulness-based interventions). Prior RCTs and systematic reviews have elucidated that psychosocial and psychological interventions, particularly those that are tailored specifically for diabetes and have a patient empowerment or motivational interviewing component, are more successful at improving glycemic outcomes in addition to reducing diabetes distress [5-9]. The exact mechanisms behind this relationship are not clear, but drawing on well-established behavioral theories may help to clarify this link. Perceived competence and self-efficacy, or the belief in an individual’s ability to complete a task, is a key feature of social cognitive theory [10], and it has been found to be consistently negatively correlated with distress and is in the mechanistic pathway between diabetes distress and self-management behaviors in T2D [11,12]. It is therefore likely that improving [2] perceived competence is an important element of interventions that improve both diabetes distress and glycemic control. Similarly, self-determination theory postulates that autonomy support, defined as the provision of social support in a way that respects the patient’s values, autonomy, and choice, is an important motivator for patients with chronic disease such as diabetes [13]. As such, autonomy support has also been shown to be an important buffer against the effects of diabetes distress on glycemic outcomes [14]. However, beyond this, there is not a consistent strategic approach common among interventions that improves both diabetes distress and glycemic control. Further elucidation is thus needed to ensure that effective intervention components that improve these constructs are incorporated into future interventions for diabetes mellitus.

Equally important is understanding the characteristics of participants who benefit the most from these interventions. Prior studies have found that patients who are younger, female, have longer duration of diabetes, and are of ethnic minority status, particularly African Americans, have higher diabetes distress levels [15-17]. Interventions targeting specific ethnic minority populations who experience disproportionate diabetes burden and elevated diabetes distress levels have shown mixed findings. These studies, however, are limited by small sample sizes and do not allow comparisons of effects across participants of different ethnicities [18]. Similarly, diabetes-specific characteristics of those who respond to interventions specifically for distress are unknown. As may be anticipated, high diabetes distress levels are associated with fear of insulin use in insulin-naive patients [19], but it is unclear whether interventions targeting distress are as effective in insulin users as in noninsulin users.

Peer support interventions, in which an individual with prior experience or knowledge who has been successful in their own self-management behaviors serves as a supportive mentor for a target population of patients with similar ethnic or socioeconomic background, are emerging as an important tool for patients with diabetes mellitus, particularly for vulnerable patient populations [14]. Peer support interventions have been successful in improving both glycemic outcomes and psychosocial outcomes, including diabetes distress, and are an attractive, low-cost approach for health care systems [20-22].

A recently published randomized controlled trial (RCT) of peer support versus technology-enhanced peer support for primarily African American veterans with T2D who receive care at an urban Veterans Affairs (VA) health center published by Heisler et al [23] demonstrated that the peer coach model they evaluated, both with and without technology enhancement, was effective at improving glycemic control and reducing diabetes distress over the 6-month intervention period.

In this trial, participants were randomized to peer coaches without any additional eHealth tools or to peer coaches using an individually tailored, web-based educational tool (iDecide) over the course of 6 months. This tool had interactive features to allow participants to understand their personal diabetes risk profile as well as explore options for medications based on cost, effectiveness, and side effects [23]. Peer coaches all received training in motivational interviewing [23]. In this trial, both arms achieved statistically and clinically significant improvements in both diabetes distress and $HbA_1c$ without any significant difference between the two intervention arms [23]. This successful trial thus presents an opportunity to explore the

(KEYWORDS diabetes mellitus; diabetes distress; health behavior; peer support

(JMIR Diabetes 2021;6(1):e21400) doi:10.2196/21400)
psychosocial mechanisms that lead to improvements in glycemic control when diabetes distress is reduced as well as the participant baseline characteristics that may predict responsiveness to such an intervention. The objectives of this study were therefore to evaluate mediators and moderators in the relationship between change in diabetes distress and change in glycemic control over a 6-month period in response to a peer support intervention.

**Methods**

**Conceptual Model for Mediator and Moderator Analysis**

A mediator analysis is one method to explore the psychosocial mechanisms that link diabetes distress and glycemic control. In such an analysis, a conceptual model is created that hypothesizes potential targets, or mediators, along the mechanistic pathway that an intervention must include in order to be successful in achieving the desired outcome. In the previously mentioned RCT by Heisler et al [23], participants had at least weekly contact with a fellow patient with T2D who had received a 2-hour training session with a focus on motivational interviewing, including active listening skills, rolling with resistance, enhancing change talk, goal setting, and action planning. During these sessions, peer coaches helped participants develop and follow up on weekly action steps to meet the participants’ defined behavioral goals. In order to ensure fidelity and help further strengthen the peer coach’s motivational interviewing skills, we held monthly hour-long booster sessions to provide reinforcement and additional training to coaches throughout the intervention period. Based on self-determination theory, which postulates that patients with diabetes who experience more autonomy supportiveness by their health care providers and supporters are more motivated and perceive themselves to be more competent in diabetes self-management, we hypothesized that both intrinsic motivation and perceived competence are important targets in the mechanistic pathway between diabetes distress and glycemic control [24]. Similarly, based on prior studies demonstrating the importance of goal setting and decisional conflict, we hypothesized that both are crucial elements of self-management support interventions to improve both diabetes distress and glycemic control [25]. Our full mediation model is demonstrated in Figure 1 with the pathway through relationship a and relationship b demonstrating the fully mediated model through our hypothesized mediators of goal setting, perceived competence, intrinsic motivation, and decisional conflict.

Figure 1. Conceptual model for hypothesized mediators and moderators of improved glycemic control in a peer coaching intervention.

A moderator analysis can be used to evaluate the characteristics of participants who benefited the most from the peer support intervention of reducing diabetes distress to improve glycemic outcomes. These characteristics are called moderators as they help inform differential effects in the relationship between an independent and dependent variable and hence identify potential modifiers and/or target population for the intervention. In our conceptual model shown in Figure 1, we hypothesized that
potential moderators include baseline patient characteristics (age, race, education, employment, and health literacy), certain diabetes characteristics (duration of diabetes, HbA1c, and insulin use), diabetes-specific social support, and comorbid depression. Our specific questions were as follows:

- In an intervention that improves both diabetes distress and glycemic control, are improvements in diabetes distress correlated with improvements in HbA1c (main effect)?
- Do goal setting, perceived competence, intrinsic motivation, and decisional conflict work individually or in combination to mediate the relationship between diabetes distress and glycemic control (mediating effect)?
- Does age, race, education, employment, health literacy, duration of diabetes, HbA1c, insulin use, diabetes-specific social support, or depression moderate the relationship between diabetes distress and glycemic control (moderating effect)?

### Setting, Recruitment, Intervention, and Measures

The target population for this study included veterans with T2D and high baseline HbA1c values enrolled in a comparative effectiveness RCT of peer support versus technology-enhanced peer support. The description of recruitment, intervention, outcomes, and results of this RCT have been described previously [23]. Glycemic control was measured using HbA1c at baseline and 6 months. Diabetes distress and potential mediators were measured using validated surveys at baseline and 6 months, which were then scaled from 0 to 100, with higher numbers indicating more positive outcomes (eg, lower diabetes distress, higher goal setting). Specifically, the following scales were used (see Multimedia Appendix 1 for further details):

- Diabetes distress: Measured, analyzed, and reported using the 2-item validated Diabetes Distress Scale–2, which assesses feelings that living with diabetes is overwhelming and/or that the participant is failing in their diabetes management [26,27].
- Goal setting: Measured, analyzed, and reported using the 3-item goal setting subscale of the Patient Assessment of Chronic Illness Care, which assesses whether participants were aided in setting goals for self-management and, if so, whether an action plan was developed [28].
- Perceived competence: Measured, analyzed, and reported using the 4-item validated Perceived Competence scale, which assesses the extent to which a participant feels confident and capable of meeting the challenges of diabetes self-management [13].
- Intrinsic motivation: Measured, analyzed, and reported using the intrinsic motivation subscale of the Treatment Self-Regulation Questionnaire, which assesses the extent to which participants feel self-motivated to improve their health behaviors [13].
- Decisional conflict: Measured, analyzed, and reported using the 1-item validated Decisional Conflict Scale, which assess the extent to which a participant is satisfied with their medication options for diabetes [29].

In the RCT, both arms demonstrated improved diabetes distress and HbA1c values at 6 months. Therefore, in this study, participants in both arms were combined to investigate goal setting, perceived competence, intrinsic motivation, and decisional conflict as potential mediators, as shown in Figure 1. Additionally, baseline characteristics were evaluated as moderators of improvement in both diabetes distress and glycemic control, also shown in Figure 1.

### Statistical Analysis

Descriptive statistics were used to evaluate frequencies and means of baseline participant characteristics, and paired t tests were used to evaluate the change in means from baseline to 6 months for the independent variable, dependent variable (HbA1c), and hypothesized mediator variables (goal setting, perceived competence, intrinsic motivation, and decisional conflict). Bivariate and multivariable linear regressions were used to assess whether the change in diabetes distress at 6 months (independent variable) is associated with the change in HbA1c at 6 months (dependent variable). Covariates include age, race, health literacy, duration of diabetes, and baseline HbA1c.

We next assessed the role of goal setting, perceived competence, intrinsic motivation, and decisional conflict as mediators between the change in diabetes distress and the change in HbA1c at 6 months. Multivariable linear regression models were used with the covariate adjustments of age, race, health literacy, duration of diabetes, and baseline HbA1c. This is conceptualized by the mediation model in Figure 1:

- Relationship a: between diabetes distress (independent variable) and all potential mediators (dependent variables)
- Relationship b: between all potential mediators (independent variable) and HbA1c

The potential mediators that were found to be significantly associated with the change in diabetes distress and HbA1c at 6 months were selected for formal mediation testing by using seemingly unrelated linear regression techniques [30]. We evaluated each individual mediator separately as well as the shared effect of the combined mediators on the mediation pathway through relationships a and b (the indirect pathway) [30]. We calculated bias-corrected 95% confidence intervals from a bootstrapping method with 5000 replications [30].

Finally, sociodemographic factors (age, race, highest attained education, income, employment) and baseline clinical and psychosocial attributes (health literacy, HbA1c, duration of diabetes, insulin use, diabetes-specific social support, depressive symptoms) were assessed as potential moderators of the relationship between change in diabetes distress and change in HbA1c at 6 months. Multivariable linear regressions include an interaction term between the change in diabetes distress at 6 months and each of the potential moderators as well as those variables themselves. The change in HbA1c at 6 months was the independent variable in these models and covariates included age, race, health literacy, duration of diabetes, and baseline HbA1c except where the variable was tested as a moderator.

This moderator model is conceptualized in Figure 1 (ie, differential effects on relationship d). For each potential moderator, the significance of the interaction term was assessed
for different subgroups, and the difference in coefficients between the subgroups was evaluated for significance.

**Results**

**Description of the Sample**

A total of 290 veterans with T2D were enrolled in the two intervention arms of the RCT. Baseline characteristics of the full cohort are shown in Table 1. Being a veteran population, 98% of the participants were male with an average age of 63 (SD 10.2) years, and 63% were African American. The average HbA$_1c$ was 9.1% (SD 1.7%) with a mean of 15 years of diabetes duration, and 60% of the participants were insulin-dependent. At 6 months, diabetes distress improved by 4.8 points (95% CI 2.2 to 7.5; \( P < .001 \)) and mean HbA$_1c$ levels improved by 0.7% (95% CI –0.9 to –0.5; \( P < .001 \)) in all participants (Multimedia Appendix 2). Scores for goal setting, perceived competence, intrinsic motivation, and decisional conflict improved by 14.3, 6.9, 6.8, and 6.8 points, respectively (all \( P < .001 \)) at 6 months (Multimedia Appendix 2).
Table 1. Baseline characteristics of all participants (n=290).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, mean (SD)</td>
<td>63 (10.2)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Male</td>
<td>283 (98)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>181 (62)</td>
</tr>
<tr>
<td>White</td>
<td>106 (37)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Work status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>74 (26)</td>
</tr>
<tr>
<td>Not employed</td>
<td>49 (17)</td>
</tr>
<tr>
<td>Retired</td>
<td>141 (49)</td>
</tr>
<tr>
<td>Disabled</td>
<td>23 (8)</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>12 (4)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>78 (27)</td>
</tr>
<tr>
<td>Some tech or vocational</td>
<td>23 (8)</td>
</tr>
<tr>
<td>Some college or more</td>
<td>177 (61)</td>
</tr>
<tr>
<td>Income ($), n (%)</td>
<td></td>
</tr>
<tr>
<td>1-15,000</td>
<td>61 (21)</td>
</tr>
<tr>
<td>16,000-30,000</td>
<td>81 (28)</td>
</tr>
<tr>
<td>31,000-55,000</td>
<td>59 (20)</td>
</tr>
<tr>
<td>56,000 and above</td>
<td>46 (16)</td>
</tr>
<tr>
<td>Prefer not to discuss</td>
<td>42 (15)</td>
</tr>
<tr>
<td>Baseline HbA1c a, mean (SD)</td>
<td>9.1 (1.7)</td>
</tr>
<tr>
<td>Number of years with diabetes, mean (SD)</td>
<td>15.2 (10.0)</td>
</tr>
<tr>
<td>Insulin use, n (%)</td>
<td>171 (60)</td>
</tr>
<tr>
<td>Number of oral antihyperglycemic meds, mean (SD)</td>
<td>1.1 (0.8)</td>
</tr>
<tr>
<td>Health literacy, mean (SD)</td>
<td>7.0 (1.9)</td>
</tr>
<tr>
<td>Diabetes-specific social support b, mean (SD)</td>
<td>54.4 (14.3)</td>
</tr>
<tr>
<td>Depression c, mean (SD)</td>
<td>76.9 (27.0)</td>
</tr>
</tbody>
</table>

aHbA1c: hemoglobin A1c.
bBased on the Diabetes-Specific Social Support Needs assessment [31], scaled score ranging from 0 to 100, with more positive outcomes reflected by higher numbers.
cBased on the Patient Health Questionnaire–2 scaled score ranging from 0 to 100, with more positive outcomes reflected by higher numbers.

Results of the Main Relationship
A significant association between the improvement in diabetes distress and decreased HbA1c was found in the unadjusted model (β-coefficient -0.017; 95% CI -0.028 to -0.006; P=.003) (relationship d). This association remained significant in the adjusted model, controlling for age, race, health literacy, duration of diabetes, and baseline HbA1c (β-coefficient -0.015; 95% CI -0.025 to -0.006; P=.001).

Results of the Mediator Analysis
Improvement in goal setting at 6 months was associated with improvements in diabetes distress (β coefficient 0.225, P=.02) and reduction in the HbA1c (β coefficient -0.009, P=.004) at 6 months. Similarly, improvement in perceived competence at 6 months was associated with both improvements in diabetes distress (β coefficient 0.182, P=.002) and the improvement in HbA1c (β coefficient -0.011, P=.03) at 6 months. Neither
intrinsic motivation or decisional conflict were associated with the change in diabetes distress or change in HbA1c at 6 months so were removed from further mediation analyses. These results are highlighted in Table 2.

Table 2. Adjusted estimates of the effect of diabetes distress on all potential mediators (relationship a) and the effect of all mediators on hemoglobin A1c (relationship b).a

<table>
<thead>
<tr>
<th>Potential mediator (outcome in relationship a; predictor in relationship b)</th>
<th>Main predictor: diabetes distressb (relationship a)</th>
<th>Main outcome: hemoglobin A1c (relationship b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β coefficient</td>
<td>95% CI</td>
</tr>
<tr>
<td>Goal setting</td>
<td>.225</td>
<td>.036 to .414</td>
</tr>
<tr>
<td>Perceived competence</td>
<td>.183</td>
<td>.065 to .300</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>.007</td>
<td>−.127 to .141</td>
</tr>
<tr>
<td>Decisional conflict</td>
<td>.101</td>
<td>−.053 to .255</td>
</tr>
</tbody>
</table>

aDiabetes distress, hemoglobin A1c, and all potential mediators assessed as the mean change from baseline to 6 months.

bModels included diabetes distress as the independent variable and potential mediators as dependent variables; covariates include age, race, health literacy, duration of diabetes, and baseline A1c variables.

cModels included potential mediators as the independent variable and hemoglobin A1c as the dependent variable; covariates include age, race, health literacy, duration of diabetes, and baseline A1c variables.

Table 3 presents the extent to which the association between improvement in HbA1c and the improvement in diabetes distress was mediated by goal setting or perceived competence (through the pathway that encompasses relationships a and b in Figure 1). We found that both goal setting and perceived competence are modest mediators with a combined 20% shared total effect (combined indirect effect −0.003, 95% CI −0.0072 to −0.0005).

Table 3. Mediating effects of goal setting and perceived competence in the relationship between diabetes distress and hemoglobin A1c (mediator analysis).

<table>
<thead>
<tr>
<th>Potential mediatora</th>
<th>Indirect effectb (95% CI)</th>
<th>Share of total effect (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setting</td>
<td>−0.002 (−0.0052 to −0.0001)</td>
<td>13.3</td>
</tr>
<tr>
<td>Perceived competence</td>
<td>−0.001 (−0.0045 to −0.0002)</td>
<td>6.7</td>
</tr>
<tr>
<td>Combination of goal setting and perceive competence</td>
<td>−0.003 (−0.0072 to −0.0005)</td>
<td>20</td>
</tr>
</tbody>
</table>

aGoal setting and perceived competence assessed as the mean change from baseline to 6 months.

bCovariates include age, race, health literacy, duration of diabetes, and baseline hemoglobin A1c.

Results of the Moderator Analysis

As shown in Table 4, the within-group estimates for the relationship between the change in diabetes distress and the change in HbA1c at 6 months was significant for participants who are younger than age 65 years, have more than a high school education, are employed, have an income greater than $30,000 per year, have lower health literacy, have more depressive symptoms, who have more social support, who have had diabetes for fewer years, and those with a baseline HbA1c <8.5%. The between group estimates suggest there is a significant difference in the relationship between the change in diabetes distress and the change in HbA1c at 6 months by race and the status of insulin use: stronger for whites compared with African Americans (P=.002) and for those who were using insulin compared with those not (P=.02).
Table 4. Adjusted estimates on the effect of improved diabetes distress on improved glycemic control, by groups with different baseline characteristics (moderator analysis).

<table>
<thead>
<tr>
<th>Potential moderator</th>
<th>N</th>
<th>Baseline mean diabetes distress (Predictor)</th>
<th>Baseline mean HBA\textsubscript{1c} (Outcome)</th>
<th>Adjusted estimates</th>
<th>( P ) value</th>
<th>Difference in ( \beta ) coefficients (between subgroups)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age in years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>154</td>
<td>71.7</td>
<td>9.3</td>
<td>-0.019</td>
<td>.002</td>
<td>0.007</td>
<td>.24</td>
</tr>
<tr>
<td>&gt;65</td>
<td>136</td>
<td>74.9</td>
<td>8.8</td>
<td>-0.012</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>181</td>
<td>74.0</td>
<td>9.1</td>
<td>-0.006</td>
<td>.28</td>
<td>0.029</td>
<td>.002</td>
</tr>
<tr>
<td>White</td>
<td>106</td>
<td>72.2</td>
<td>9.0</td>
<td>-0.035</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;HS\textsuperscript{c}</td>
<td>12</td>
<td>77.8</td>
<td>8.8</td>
<td>0.024</td>
<td>.52</td>
<td>0.040</td>
<td>.63</td>
</tr>
<tr>
<td>&gt;HS</td>
<td>278</td>
<td>73.0</td>
<td>9.1</td>
<td>-0.016</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None\textsuperscript{d}</td>
<td>213</td>
<td>74.6</td>
<td>9.1</td>
<td>-0.011</td>
<td>.19</td>
<td>0.008</td>
<td>.58</td>
</tr>
<tr>
<td>Employed</td>
<td>74</td>
<td>69.6</td>
<td>8.9</td>
<td>-0.018</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Income ($)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30,000</td>
<td>142</td>
<td>73.1</td>
<td>9.1</td>
<td>-0.012</td>
<td>.07</td>
<td>0.011</td>
<td>.13</td>
</tr>
<tr>
<td>&gt;30,000</td>
<td>105</td>
<td>73.8</td>
<td>9.0</td>
<td>-0.023</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health literacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>152</td>
<td>70.4</td>
<td>9.1</td>
<td>-0.026</td>
<td>&lt;.001</td>
<td>0.018</td>
<td>.07</td>
</tr>
<tr>
<td>High</td>
<td>138</td>
<td>76.3</td>
<td>9.1</td>
<td>-0.008</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline depression\textsuperscript{e}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>132</td>
<td>81.9</td>
<td>8.8</td>
<td>-0.013</td>
<td>.10</td>
<td>0.003</td>
<td>.64</td>
</tr>
<tr>
<td>High</td>
<td>158</td>
<td>66.0</td>
<td>9.3</td>
<td>-0.015</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline social support\textsuperscript{f}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>111</td>
<td>76.9</td>
<td>9.2</td>
<td>-0.012</td>
<td>.15</td>
<td>-0.004</td>
<td>.59</td>
</tr>
<tr>
<td>High</td>
<td>130</td>
<td>72.2</td>
<td>9.0</td>
<td>-0.016</td>
<td>.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Duration of diabetes in years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>111</td>
<td>71.4</td>
<td>9.3</td>
<td>-0.026</td>
<td>.006</td>
<td>0.016</td>
<td>.05</td>
</tr>
<tr>
<td>&gt;10</td>
<td>179</td>
<td>74.3</td>
<td>8.9</td>
<td>-0.008</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline HBA\textsubscript{1c} (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;8.5</td>
<td>109</td>
<td>78.1</td>
<td>7.7</td>
<td>-0.021</td>
<td>.004</td>
<td>0.011</td>
<td>.50</td>
</tr>
<tr>
<td>&gt;8.5</td>
<td>134</td>
<td>70.8</td>
<td>10.2</td>
<td>-0.010</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Insulin use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>119</td>
<td>73.7</td>
<td>8.8</td>
<td>-0.006</td>
<td>.40</td>
<td>0.024</td>
<td>.02</td>
</tr>
<tr>
<td>Yes</td>
<td>171</td>
<td>72.9</td>
<td>9.3</td>
<td>-0.029</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^{a} \text{HBA}_{1c}: \) hemoglobin A\textsubscript{1c}.

\( ^{b} \text{Adjusted for age, race, health literacy, duration of diabetes and baseline hemoglobin A}\textsubscript{1c} except where these variables were tested as moderators. \)

\( ^{c} \text{HS: high school.} \)

\( ^{d} \text{Includes not employed, retired and disabled.} \)
improvement of the HbA1c due to the reduced levels of diabetes distress. Race was found to a moderator, suggesting that Caucasian veterans responded more to the peer support intervention than African American patients. Prior studies suggest that peer supporters who are culturally appropriate (including concordant age, race, and gender) may be more effective peer supporters for African Americans with diabetes [34,35]. Given that the burden of T2D falls heavily on minority populations, including African American and Latino populations [36], further studies are needed to understand the characteristics of effective interventions that target these high-risk populations, such as cultural concordance among peer supporters. Additionally, insulin use was found to be a moderator, suggesting that peer support interventions targeting high distress levels in insulin-requiring T2D patients lead to better glycemic control. This is important because approximately one-quarter of T2D patients in the United States currently require insulin, and this proportion is on the rise [37].

Strengths and Limitations
This study has several strengths. The first is that, to our knowledge, this is the first study looking at mediators and moderators between glycemic control and diabetes distress in an intervention that improves both. We incorporated robust statistical methods to assess the mediation pathway, finding that goal setting and perceived competence are important for future interventions targeting both glycemic and psychosocial outcomes for T2D. This is also one of the first studies to more specifically examine a broad array of socioeconomic and diabetes-specific characteristics that might moderate the relationship between diabetes distress and glycemic control. This is important because this can facilitate screening and targeted interventions using information readily captured by electronic medical records.

We also recognize that our study has several important limitations. First, this study was conducted in primarily African American male veterans with T2D, which limits the generalizability of our findings. It is therefore possible that, in other populations, goal setting and perceived competence have less significance in the mechanistic pathway between elevated levels of diabetes distress and worse glycemic control. Additionally, our use of brief validated scales to measure multiple complicated psychological constructs is a potential limitation, as these short-form scales did not permit in-depth investigation into different facets of these constructs. For example, we used the Diabetes Distress Scale 2 to measure diabetes distress, rather than the full 17-item Diabetes Distress Scale. Although the 2-item Diabetes Distress Scale has been found to correlate well with the larger Diabetes Distress Scale questionnaire, it does not provide subtypes of distress as it only measures emotional distress and this may have impacted our moderator analyses [27]. Prior studies indicate Black patients have higher levels of provider-related distress [38], which was not specifically measured in our study. It is possible that there are differences in the subtypes of diabetes distress (emotional burden, provider-related, interpersonal, and regimen-related).
among different populations (such as race/ethnicity) that account for the differential response in White versus Black participants in our study. The study population was also nearly exclusively male and does not therefore generalize to women with T2D, who often have higher levels of diabetes distress [39]. Future studies should include evaluation of interventions of women with T2D with high diabetes distress levels and use of more comprehensive scales to measure diabetes distress in order to more accurately generalize to all T2D populations. Finally, we hypothesized a priori that there would be 4 potential mediators and found that only goal setting and perceived competence were mediators. However, combined, these mediators only accounted for 20% of the mediation effect, suggesting that there are other important mediators in the mechanistic pathway between diabetes distress and glycemic control that we did not measure. Future studies are therefore needed to clarify these additional mediating mechanisms.

**Conclusion**

In conclusion, we found that in a peer support intervention for T2D in primarily African American male veterans both goal setting and perceived competence are important mediators in the mechanistic pathway between diabetes distress and glycemic control. Additionally, we found that this peer support intervention that improved diabetes distress was most effective in reducing HbA$_1c$ levels in White and insulin-requiring veterans with T2D. These findings are important for informing future interventions that target both psychosocial and glycemic outcomes and efforts to tailor interventions to best meet the needs of patients with different characteristics.

**Acknowledgments**

This research was supported by grants from the Veterans Affairs Health Services Research and Development Service (12-412) and the National Institute of Diabetes and Digestive and Kidney Diseases (P30DK092926 MCDTR).

**Authors' Contributions**

KMS, HC, GP, and MH designed the study. HC and MH collected the data. KMS, HC, and CR analyzed the data. KMS wrote the first draft of the manuscript. KMS, HC, CR, GP, and MH edited the manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Diabetes distress, goal setting, perceived competence, intrinsic motivation, and decisional conflict scales.

[DOCX File, 204 KB - diabetes_v6i1e21400_app1.docx]

**Multimedia Appendix 2**

Summary of the change in diabetes distress, change in HbA1c, and hypothesized mediators between baseline and 6 months.

[DOCX File, 14 KB - diabetes_v6i1e21400_app2.docx]

**References**


Abbreviations

HBA1c: hemoglobin A1c
RCT: randomized controlled trial
T2D: type 2 diabetes
VA: Veterans Affairs

©Kara Mizokami-Stout, Hwajung Choi, Caroline R Richardson, Gretchen Piatt, Michele Heisler. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 11.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.
Using Wearable Activity Trackers to Predict Type 2 Diabetes: Machine Learning–Based Cross-sectional Study of the UK Biobank Accelerometer Cohort

Benjamin Lam1, BSc; Michael Catt2, PhD; Sophie Cassidy3, PhD; Jaume Bacardit1, PhD; Philip Darke1, BSc; Sam Butterfield1, BSc; Ossama Alshabrawy1, PhD; Michael Trenell4, PhD; Paolo Missier1, PhD

1School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom
2Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom
3Faculty of Medicine and Health, University of Sydney, Sydney, Australia
4Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom
5Faculty of Medical Sciences, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Benjamin Lam, BSc
School of Computing
Newcastle University
Urban Sciences Building
1 Science Square
Newcastle upon Tyne, NE4 5TG
United Kingdom
Phone: 44 7704111910
Email: b.p.lam1@ncl.ac.uk

Abstract

Background: Between 2013 and 2015, the UK Biobank collected accelerometer traces from 103,712 volunteers aged between 40 and 69 years using wrist-worn triaxial accelerometers for 1 week. This data set has been used in the past to verify that individuals with chronic diseases exhibit reduced activity levels compared with healthy populations. However, the data set is likely to be noisy, as the devices were allocated to participants without a set of inclusion criteria, and the traces reflect free-living conditions.

Objective: This study aims to determine the extent to which accelerometer traces can be used to distinguish individuals with type 2 diabetes (T2D) from normoglycemic controls and to quantify their limitations.

Methods: Machine learning classifiers were trained using different feature sets to segregate individuals with T2D from normoglycemic individuals. Multiple criteria, based on a combination of self-assessment UK Biobank variables and primary care health records linked to UK Biobank participants, were used to identify 3103 individuals with T2D in this population. The remaining nondiabetic 19,852 participants were further scored on their physical activity impairment severity based on other conditions found in their primary care data, and those deemed likely physically impaired at the time were excluded. Physical activity features were first extracted from the raw accelerometer traces data set for each participant using an algorithm that extends the previously developed Biobank Accelerometry Analysis toolkit from Oxford University. These features were complemented by a selected collection of sociodemographic and lifestyle features available from UK Biobank.

Results: We tested 3 types of classifiers, with an area under the receiver operating characteristic curve (AUC) close to 0.86 (95% CI 0.85-0.87) for all 3 classifiers and F1 scores in the range of 0.80-0.82 for T2D-positive individuals and 0.73-0.74 for T2D-negative controls. Results obtained using nonphysically impaired controls were compared with highly physically impaired controls to test the hypothesis that nondiabetic conditions reduce classifier performance. Models built using a training set that included highly impaired controls with other conditions had worse performance (AUC 0.75-0.77; 95% CI 0.74-0.78; F1 scores in the range of 0.76-0.77 for T2D positives and 0.63-0.65 for controls).

Conclusions: Granular measures of free-living physical activity can be used to successfully train machine learning models that are able to discriminate between individuals with T2D and normoglycemic controls, although with limitations because of the intrinsic noise in the data sets. From a broader clinical perspective, these findings motivate further research into the use of physical activity as a predictor of T2D.
Introduction

The UK Biobank

Objective measures of physical activity can be used to characterize people’s free-living movement behavior to provide the kind of digital phenotype [1] that promises to support a vision of participatory, preventive, and personalized health care. The UK Biobank collected the largest available data set of free-living physical activity traces [2]. It includes uncontrolled, raw accelerometry traces collected for 7 days for a random selection of 103,712 out of a total of 502,664 UK Biobank participants (approximately 25%) between February 2013 and December 2015. All the studies cited here, including the one described in this paper, have used a reduced set after performing quality checks.

This data set has been used in recent studies to quantify differences in physical activity levels across the general UK Biobank population [3] and to show that participants with chronic diseases exhibit lower levels of activity than the general UK Biobank cohort [4]. It has also demonstrated associations between cardiometabolic health, multimorbidity, and mortality [5,6]. However, this data set has not been used to validate the hypothesis that accelerometer traces measures of physical activity can be used as a predictor for type 2 diabetes (T2D) and, thus, potentially, as a valid digital phenotype for early detection of T2D.

T2D and Physical Activity

T2D is linked with low physical activity levels and increasing age [7]. This disease has become much more prevalent and is rapidly rising globally, especially in parts of the developing world [8].

Research into the effectiveness of activity monitoring for T2D detection and prevention is motivated by the disproportionately high cost, both economic and social, of treating T2D [9], considering that approximately 90%-95% of diagnosed diabetes among adults is type 2. In the United Kingdom alone, more than 2.7 million people have been diagnosed with T2D, whereas a further 750,000 people are believed to have the symptoms but are yet to be diagnosed with the disease [10].

Studies have been undertaken to use digital phenotypes for early diagnosis, but most studies have focused on using traditional multi-omics approaches [11].

The UK Biobank Accelerometer Data and T2D

In this study, we tested the hypothesis that activity profiles, when represented in sufficient detail, differ significantly between individuals with T2D and the general population.

This study begins by defining participants with T2D in the UK Biobank using a combination of preexisting diagnoses collected in the UK Biobank assessment centers and automated analysis of the participants’ electronic health records (EHRs) follow-up. We then evaluate the extent to which accelerometer traces can distinguish individuals with T2D from normoglycemic controls. The approach employs a combination of traditional machine learning classification models to quantify the predictive power of features extracted from accelerometer traces and to assess their limitations relative to this task.

Methods

Overview

This paper refers to each volunteer’s 1-week activity recording period as their wear time and to the UK Biobank volunteers as the accelerometer cohort.

The data set used in this study was derived from the collection of activity traces for each of these participants, filtered using the inclusion and exclusion criteria described below. Variables representing physical activity features were extracted from the raw traces. In addition, a small set of sociodemographic, anthropometric, and metabolic variables were added, following recent studies [11] in which the same variables were used to characterize the behavioral phenotype of UK Biobank participants relative to cardiovascular disease (CVD) and T2D.

Inclusion and Exclusion Criteria for T2D-Positive Participants

The criteria described below and the resulting data set sizes are summarized in Figure 1. Participants with T2D were identified using a combination of self-reported data collected at the Biobank assessment center and data from the participants’ primary care EHR, including prescriptions. At the time of writing, EHR records were available for approximately 245,000 out of 502,664 individuals (approximately 45%) of the UK Biobank population. Inclusion in the T2D group, based on self-reporting, follows the same criteria as in the study by Schißler-Fiorenza Rose et al [11], namely, individuals with an explicit diagnosis as part of their assessment, based on the UK Biobank Showcase [12].
Figure 1. Training set selection criteria for type 2 diabetes-negative and type 2 diabetes-positive individuals. EHR: electronic health record; QC: quality control.

At the baseline assessment center, participants who had been diagnosed with diabetes or T2D were selected; those taking insulin within their first year (variable 2986-0.0) and who were less than 35 years old (variable 2976-0.0) at diagnosis were excluded to reduce the likelihood of individuals with type 1 diabetes and monogenic forms of diabetes [13]. This resulted in 2755 participants from the accelerometry cohort being identified as having T2D.

Primary care EHRs were also used to identify participants who developed T2D after their baseline assessment but before their accelerometer wear time. The incidence of T2D was defined as the occurrence of a Read Code version 2 or Clinical Terms Version 3 (CTV3) code corresponding to T2D after the date of the assessment center visit. Read Code version 2 code sets developed by Kuan et al were used [14], and equivalent CTV3 codes were mapped using mapping data provided by the UK Biobank [4,5].

The low prevalence of T2D in the UK Biobank population is reflected in the very small positive group, compared with an overwhelmingly large non-T2D control group (99,636 participants). Therefore, it is necessary to rebalance these classes before model learning. Rather than random selection from the control group, better selection criteria can be adopted.

We observed that the normoglycemic control group might include individuals with nondiabetes-related physical activity impairments. Excluding such individuals is desirable, as it is likely to remove noise from the control group. The controls’ selection process described below includes a judgment, grounded in general medical knowledge, of how a wide variety of conditions may have affected a participant’s ability to perform...
normal activities. Although the assessment may not be entirely accurate, to the best of our knowledge, this is the first attempt to select a control group based on EHR data. The outcome was assumed to be no worse than random selection from the control group. The results show that the prediction accuracy improves relative to using a random control training set.

The selection process involved a further analysis of EHRs for a period antecedent of wear time to identify any nondiabetes medical conditions that may have resulted in physical activity impairment. This analysis is limited by the partial availability of EHRs (approximately 20,000 individuals within the cohort). The analysis is described in detail in Multimedia Appendix 1.

An impairment score is calculated for each individual by (1) associating a severity score with each type of relevant disease reference in the Read Code version 2 catalog and (2) averaging the scores across all occurrences of the disease references in the individual’s EHR history, within 6 months before wear time. Records are included for 1 month after wear time, as there may be a delay in recording new conditions. The analysis resulted in 2 control subpopulations, as shown in Figure 1 (bottom right): Norm-0, where we expected no impairment (n=8463), and Norm-2, with expected high impairment (n=1666). These results are summarized in Table 1. Both sets were used as part of supervised learning in separate experiments, as explained below.

### Table 1. Number of participants in each subpopulation according to activity impairment severity score.

<table>
<thead>
<tr>
<th>Impairment score</th>
<th>Total participants, N</th>
<th>Participants with adequate wear time, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm-0</td>
<td>11,019</td>
<td>8463 (76.80)</td>
</tr>
<tr>
<td>Norm-2</td>
<td>3355</td>
<td>1666 (49.66)</td>
</tr>
</tbody>
</table>

It is also acknowledged that 151 out of 3101 T2D-positive individuals also had a high impairment severity score for physical activity. This small subset of the T2D-positive population was not excluded from the training data sets. T2D is a complex disease that can cause many complications or comorbidity with other conditions, such as CVD. Therefore, to capture all behaviors and activity patterns associated with T2D, it is important to include the severely impaired T2D-positive individuals in the overall T2D-positive population.

We have also experimentally verified that removing these few individuals from the training set does not alter the properties of the resulting model (refer to the Results section).

### Training Data Sets

Using these 2 control groups, 2 training sets were formed: training set 1: T2D versus Norm-0 and training set 2: T2D versus Norm-2. The first was used to test our main hypothesis that activity levels in the T2D group were significantly different from those in the unimpaired control group. The second was used to quantify the effect of possible nondiabetic activity impairment as a source of noise in the controls. This was achieved by training the same models using training set 1 and training set 2 and then comparing their relative predictive performance.

### Physical Activity Features

A raw accelerometry trace consists of a triaxial (x, y, and z) time series. The open-source accelerometer analysis toolkit developed at the University of Oxford, available on GitHub [15], was used to annotate timelines for each raw activity trace [16]. The tool breaks down the time series into 30-second fragments, called epochs, and then employs a classifier (random forests and hidden Markov models) to annotate a time series in which each epoch belongs to 1 of 5 activity types: sedentary, moderate, walking, sleep, and light tasks. This tool distinguishes between walking from sedentary and moderate activities. According to the authors of this study, these activity types correspond to the following metabolic equivalent of task levels: sedentary, 1.5; moderate, 4.9; walking, 3.2; sleep, 1.0; and light tasks, 2.2. The feature extraction hierarchy is summarized in Figure 2.
The level above the time series activity recognition sequence uses activity bouts. An activity bout is defined as a single epoch or an uninterrupted consecutive series of epochs in which a single activity type is performed. The length of a bout refers to the many 30-second epochs for which each bout is performed. The features extracted for this study are at the level of activity bouts of each activity type: their frequency, average length, and percentage of time spent in each, broken down into fractions of a 24-hour day. This choice is inspired by neuroscience research on the effects of cognitive impairment in early stages of Parkinson disease on gait, where ambulatory bouts play a key role [17,18]. A personalized analysis of daily activities was performed to extract these features. First, to accommodate for different sleeping habits, night-sleep time boundaries were identified for each individual. These are defined as the average of the largest nearly continuous period of sleep activity bouts over a 24-hour period. The remaining period of the 24-hour day is then divided into 3 phases, denoted as morning, afternoon, and evening. Within each phase, the activity bout level features were extracted for each activity type.

This analysis results in a breakdown of 60 activity bout-level features, organized into a 5x4 matrix for each individual, with features extracted for four periods of the 24-hour day including sleep time as shown in Figure 3. Each element in the matrix (the type of activity and time of day) has 3 features: (1) number of bouts for that activity, (2) percentage of time spent in the activity, and (3) average length of the bouts. This arrangement resulted in a total of 60 features per individual. These were then aggregated over 7 days of wear time, taking the average for each element in the matrix. This feature space is referred to as the high-level activity bout features in this study. The code is available on GitHub [19].

Figure 3. Feature matrix for physical activity bout representation space.

Sociodemographic, Anthropometric, and Lifestyle Features

To quantify the relative importance of the new high-level activity bout features when used in machine learning, traditional sociodemographic and lifestyle indicators that are commonly associated with the incidence of T2D have been added. These are shown in Table 2 and were chosen based on previous studies [5,20]. These features are combined with self-reported physical activity assessments, some of which are not part of the output from the Oxford accelerometer analysis tool, notably vigorous activity. In contrast, the physical activity features in our
approach are the high-level activity bout features obtained from objective accelerometer measurements. Objective physical activity metrics also help to validate subjective measurements [21,22].

Table 2. Sociodemographic, lifestyle, and anthropometric characteristics selected from the UK Biobank baseline assessment for comparison with high-level activity bout features space.

<table>
<thead>
<tr>
<th>Sociodemographic, lifestyle, and anthropometry characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male or female (approximately 50:50 ratio)</td>
</tr>
<tr>
<td>Age at the assessment center</td>
<td>Recruits at baseline were aged between 40 and 69 years</td>
</tr>
<tr>
<td>Ethnic group</td>
<td>Predominantly White British, with some participants identifying as Black, Asian and Minority Ethnic groups</td>
</tr>
<tr>
<td>Alcohol drinking status</td>
<td>Participant reports if they were alcohol drinkers in the past, were currently drinking alcohol, or never had drunk alcohol</td>
</tr>
<tr>
<td>Smoking status</td>
<td>Participants report if they had smoked in the past, were currently smoking, or had never smoked</td>
</tr>
<tr>
<td>Body fat percentage</td>
<td>Percentage of fat in total body mass (a better indicator for obesity than BMI)</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>Measurement taken around the abdomen at the level of the umbilicus (belly button)</td>
</tr>
<tr>
<td>Sleep duration</td>
<td>Self-reported average duration of sleep in a day</td>
</tr>
<tr>
<td>Time spent watching television</td>
<td>Self-reported average time spent watching television per day</td>
</tr>
<tr>
<td>Townsend index</td>
<td>Metric for material deprivation within a population</td>
</tr>
<tr>
<td>Duration of walking activity</td>
<td>Self-reported average duration of time spent walking in a day</td>
</tr>
<tr>
<td>Duration of vigorous activity</td>
<td>Self-reported average duration of time spent performing vigorous activities during the day</td>
</tr>
<tr>
<td>Duration of moderate activity</td>
<td>Self-reported average duration of time spent performing moderate activity during the day</td>
</tr>
</tbody>
</table>

The International Physical Activity Questionnaire-Short Form was used for the variables measuring physical activity (including moderate, vigorous, and walking), television viewing times, and sleep duration (Table 2). Some of these sociodemographic and lifestyle features contained missing data. This was solved using a k-nearest neighbor imputer in scikit-learn [23], which calculates the missing value using the mean of k-nearest neighbors found in the training data using Euclidean distances, thus preserving the distribution of the original data.

Binary Classification

This exercise compares a number of classification models, obtained using different learning algorithms and using training sets training set 1 and training set 2, introduced earlier, in separate sets of experiments. Furthermore, different combinations of features were considered for each of the training sets: (1) high-level activity bout features only, (2) sociodemographic and lifestyle features only, and (3) high-level activity bout features combined with sociodemographic and lifestyle features.

These combinations produce a space of 6 data sets on which the models are trained. Three learning algorithms were tested on these data sets: random forest, logistic regression, and Extreme Gradient Boosting (XGBoost) algorithm. XGBoost is a relatively recent and perhaps less known algorithm [24], which has come to prominence owing to its superior performance, both in terms of training time and prediction accuracy, compared with random forests. XGBoost uses gradient boosting, an ensemble method that builds a stronger classifier by adding weaker models on top of each, iteratively, until the training data achieve a good level of prediction performance.

A total of 18 classifier models were trained using these combinations of 6 data sets and 3 algorithms. A standard 10-fold cross-validation was used to avoid overfitting. When learning the classifiers, a random selection of half the Norm-0 T2D-negative controls in training set 1 only was undertaken to balance the size of the Norm-0 T2D-negatives and T2D-positive (3103 individuals). Norm-0 T2D-negative individuals still vastly outnumbered the T2D-positive population.

Following common practice for binary classifiers, this study reports F1 scores, precision, recall, and area under the receiver operating characteristic curve (AUC) scores. F1 conveys the balance between precision and recall and is a value between 0 and 1, where 1 indicates perfect precision and recall. It is calculated using the harmonic mean of the precision and recall. The AUC is a metric, with values between 0 and 1, for how well a classifier is capable of distinguishing between 2 classes. A value of 1 implies a good measure of discrimination, whereas a value of 0.5 implies no discrimination capacity.

On the basis of these performance and evaluation metrics, models were compared to assess (1) the differences in predictive power between the 2 feature sets using training set 1; (2) the effect of noise in controls, using training set 2; and (3) the best modeling algorithms.
Clustering Analysis

Further analysis was undertaken where unsupervised clustering algorithms were used to segregate and identify unlabeled individuals that exhibit similar behavior with the new high-level activity bout feature space. These clusters were then profiled and interpreted in terms of their anthropometric, lifestyle, and sociodemographic characteristics. This analysis is beyond the scope of this paper but is reported in Multimedia Appendix 2.

Results

Distribution of Physical Activity Features

To summarize the distribution of the T2D-positive and Norm-0 T2D-negative populations, the high-level activity bout features were aggregated for a 24-hour period and averaged across both populations.

On average, both the T2D-positive and T2D-negative populations do not undertake significantly different quantities of each activity type aggregated to the level of the 24-hour day with approximately 5% moderate activity, 42% sedentary activity, 38% asleep, 5% light tasks, and 10% walking. However, the high-level activity bout features also offer an insight into the regularity and length of activity bouts. The values for these features do offer some discrimination between the T2D-positive and Norm-0 T2D-negative populations. The histograms below demonstrate an example of this by showing the distribution of daily averages for bout length, the number of bouts, and the percentage of times spent on sleep activity.

The histograms in Figures 4-6 show noticeable differences between the 2 populations in the features that we have developed, when aggregated out to a day. Breaking the daily patterns into 4 distinct times of day (morning, afternoon, evening, and during sleep) would further demonstrate the differences in activity bout patterns for the 2 populations by virtue of the granularity. The combined effect of all these granular-level activity bout features produces high model accuracy, as reported below.

![Figure 4](https://diabetes.jmir.org/2021/1/e23364)

**Figure 4.** Histogram for daily average percentage times spent asleep.

![Figure 5](https://diabetes.jmir.org/2021/1/e23364)

**Figure 5.** Histogram for daily average length of sleep bouts.
Figure 6. Histogram for daily average number of sleep bouts.

Binary Classification

A summary and performance comparison across the 18 models built for this study is presented in Tables 3 and 4, where AUC measures are obtained by averaging over 10 models using cross-validation for robustness. The receiver operating characteristic (ROC) curves and AUC scores are shown in Figures 7-12. All models were split between training and test data sets with an 80:20 ratio. More detailed metrics for precision, recall, F1, and ROC curves, using 10-fold cross-validation, are available in Multimedia Appendix 3.

Table 3. Classification results measured using area under the receiver operating characteristic curve scores, showing the effect of choice of type 2 diabetes-negatives, Norm-0 (no physical activity impairment) versus Norm-2 (severe physical activity impairment). The values in the cells represent area under the receiver operating characteristic curve scores.

<table>
<thead>
<tr>
<th>Predictive model</th>
<th>High-level activity bout features</th>
<th>Sociodemographic and lifestyle</th>
<th>High-level activity bout features+sociodemographic and lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Norm-0</td>
<td>Norm-2</td>
<td>Norm-0</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.80</td>
<td>0.68</td>
<td>0.83</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.79</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td>Extreme gradient boosting</td>
<td>0.78</td>
<td>0.66</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 4. Classification results measured using F1, showing the effect of choice of type 2 diabetes-negatives, Norm-0 (no physical activity impairment) versus Norm-2 (severe physical activity impairment). The values in the cells represent F1 scores.

<table>
<thead>
<tr>
<th>Predictive model model and T2D status</th>
<th>High-level activity bout features</th>
<th>Sociodemographic and lifestyle</th>
<th>High-level activity bout features+sociodemographic and lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Norm-0</td>
<td>Norm-2</td>
<td>Norm-0</td>
</tr>
<tr>
<td>Random forest T2D-positive</td>
<td>0.65</td>
<td>0.70</td>
<td>0.65</td>
</tr>
<tr>
<td>Random forest T2D-negative</td>
<td>0.78</td>
<td>0.54</td>
<td>0.78</td>
</tr>
<tr>
<td>Logistic regression T2D-positive</td>
<td>0.66</td>
<td>0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>Logistic regression T2D-negative</td>
<td>0.77</td>
<td>0.54</td>
<td>0.79</td>
</tr>
<tr>
<td>Extreme gradient boosting T2D-positive</td>
<td>0.66</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>Extreme gradient boosting T2D-negative</td>
<td>0.77</td>
<td>0.52</td>
<td>0.76</td>
</tr>
</tbody>
</table>

T2D: type 2 diabetes.
**Figure 7.** Receiver operating characteristic curve and area under the receiver operating characteristic curve for type 2 diabetes vs Norm-0: High-level activity bout features & sociodemographic and lifestyle features combined. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic curve; T2D: type 2 diabetes.

**Figure 8.** Receiver operating characteristic curve and area under the receiver operating characteristic curve for type 2 diabetes vs Norm-0: High-level activity bout features only. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic curve; T2D: type 2 diabetes.
Figure 9. Receiver operating characteristic curve and area under the receiver operating characteristic curve for type 2 diabetes vs Norm-0: Sociodemographic and lifestyle features only. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic curve; T2D: type 2 diabetes.

Figure 10. Receiver operating characteristic curve and area under the receiver operating characteristic curve for type 2 diabetes vs Norm-2: High-level activity bout features & sociodemographic and lifestyle features combined. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic curve; T2D: type 2 diabetes.
When performance is measured using AUC, stronger results are achieved when using high-level activity bout features and sociodemographic and lifestyle in combination, as expected. Using high-level activity bout features on their own reduces performance (approximately 7%-8%). However, high-level activity bout features provide almost the same performance as traditional sociodemographic and lifestyle features on their own.
Models were also generated using alternate training data sets, where 151 T2D-positive individuals with high physical activity impairment severity scores were excluded. These models exhibit very similar performance to those presented above, suggesting that physically impaired (Norm-2) T2D-positive individuals can be used as part of the T2D positives in the training set. F1 measures in Multimedia Appendix 3 reveal differences in classification accuracy between T2D against Norm-0 controls, and T2D against Norm-2 controls. When using Norm-0 controls, negatives are more accurately predicted than T2D, presumably because of class imbalance (4178 vs 3103). It is also clear that excluding physically impaired negatives improves the results.

When Norm-2 is used, however, T2D is more accurately predicted than negatives, perhaps because in this case, Norm-2 is the minority class (1666 vs 3103) and because of potential diversity within the highly impaired control population. This will be investigated in a future study.

In all cases, the combination of high-level activity bout features and sociodemographic and lifestyle variables gives better results than using either set of features on their own, as expected. The performances of both feature sets are largely independent of the choice of the learning algorithm, as seen by the overlapping ROC curves.

Discussion

Principal Findings

Using data from the UK Biobank, this study supports the hypothesis that individuals with diagnosed T2D exhibit physical activity patterns that are significantly different from those of normoglycemic controls, thus providing novel ways to detect T2D, that is, through appropriate analysis of physical activity patterns. Although most previous studies, particularly using UK Biobank, are limited to self-reported physical activity levels [5,11,25], here we have demonstrated the benefits of extracting a more objective and granular representation of physical activity from raw accelerometry traces data, namely, by activity type and time of day or sleep time. Using these features, either on their own or in combination with a selected set of sociodemographic, anthropometric, and lifestyle variables, we have shown that appropriately trained machine learning models were able to discriminate between the 2 cohorts with good predictive accuracy.

Practical Significance

These findings suggest that it may be possible to use continuous or periodic self-monitoring of individuals at risk of T2D, specifically those in a prediabetes state, for screening and early detection of disease progression. This is particularly important as evidence shows that reversal of T2D is possible, with a higher success rate when interventions are undertaken within the first 5 years of the disease [26-28].

However, early detection is still an unsolved problem, with recent figures reporting that over 190 million people worldwide live with undiagnosed diabetes [29]. Risk scores that are routinely used for screening, such as the Leicester score, are easy to obtain but not very accurate [30].

This suggests that self-monitoring of physical activity patterns, such as those presented in this study, may complement risk scores to help with the early detection of T2D, especially in high-risk individuals. Today, this can be achieved at a low cost using readily available technology [31], including internet-enabled data loggers that do not require participants to return devices, such as smartphones, periodically. However, further research is required to establish the quality and significance of physical activity data for this specific purpose.

Limitations

In principle, it may be possible to try and detect early signs of T2D using specific fingerprint patterns found in physical activity traces, where an example of a pattern may be a person who takes short bouts of low or moderate activities with frequent sedentary breaks in between. However, in practice, we found no evidence in the UK Biobank data set that strong correlations exist between specific physical activity patterns and T2D. Thus, what the machine learning approach has to offer may be limited to the strong indication demonstrated in this work, namely, that granular features extracted from the raw traces, taken together, are indeed good predictors and usefully augment the more traditional sociodemographic set of variables.

Although the UK Biobank is the largest known public accelerometry data set where a T2D cohort can be identified, detecting differences between T2D and controls remains challenging because of their low prevalence in the population, which is reflected in this study with the relatively small data set available for training when using supervised machine learning. Simultaneously, this data set was subject to noise for two reasons. First, because no formal quality assurance protocol was enforced during data collection, and second, because of the limited knowledge about other non-T2D–related conditions among the controls, which may contribute to reduced physical mobility or a more sedentary routine. We have shown how EHRs can be used to overcome this limitation.

Conclusions

This study motivates further research into the use of granular physical activity measures as a form of digital phenotype for T2D. It also suggests that more rigorous protocols on wearing physical activity loggers are required to improve the quality of the data and the signal-to-noise ratio, along with stringent inclusion and exclusion criteria or at least comprehensive knowledge of clinical conditions that may affect the signal in the traces. This is also reflected in other studies [32,33]. When such quality criteria are met, it should be possible to repeat the analysis presented here using data sets from large-scale deployment of physical activity loggers to validate the hypothesis that early detection of T2D is scientifically and technically feasible.
Acknowledgments

The authors would like to thank all the participants and data collectors of the UK Biobank for providing the data sets which made this study possible. The authors would also like to thank Dr Doherty and his collaborators at Oxford University for making the accelerometer data analysis software libraries available in the public domain.

Authors’ Contributions

BL and PM conceived the study and wrote the manuscript. BL developed and implemented the analysis. PD and SB developed the training set inclusion and exclusion criteria. SC and MC reviewed and edited the manuscript. MT, MC, and SC were mostly responsible for making access to the UK Biobank possible.

Conflicts of Interest

MT is CEO and shareholder of Changing Health Ltd, a digital behavior change company.

Multimedia Appendix 1
Activity impairment scoring details.
[PDF File (Adobe PDF File), 115 KB - diabetes_v6i1e23364_app1.pdf]

Multimedia Appendix 2
Analysis details and results of unsupervised clustering work.
[PDF File (Adobe PDF File), 667 KB - diabetes_v6i1e23364_app2.pdf]

Multimedia Appendix 3
Full results details of analysis.
[PDF File (Adobe PDF File), 1373 KB - diabetes_v6i1e23364_app3.pdf]

References

2. URL: https://www.ukbiobank.ac.uk/ [accessed 2018-09-15]
10. Diabetes.co.uk. URL: https://www.diabetes.co.uk/type2-diabetes.html [accessed 2020-08-20]


**Abbreviations**

AUC: area under the receiver operating characteristic curve
Using Wearable Activity Trackers to Predict Type 2 Diabetes: Machine Learning–Based Cross-sectional Study of the UK Biobank Accelerometer Cohort


JMIR Diabetes 2021;6(1):e23364
URL: https://diabetes.jmir.org/2021/1/e23364
doi:10.2196/23364
PMID:33739298

©Benjamin Lam, Michael Catt, Sophie Cassidy, Jaume Bacardit, Philip Darke, Sam Butterfield, Ossama Alshabrawy, Michael Trenell, Paolo Missier. Originally published in JMIR Diabetes (http://diabetes.jmir.org), 19.03.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on http://diabetes.jmir.org/, as well as this copyright and license information must be included.