Algorithm-Enabled, Personalized Glucose Management for Type 1 Diabetes at the Population Scale: Prospective Evaluation in Clinical Practice

David Scheinker¹,²,³,⁴, PhD; Angela Gu³, MS; Joshua Grossman³, MS; Andrew Ward⁵, PhD; Oseas Ayerdi³, MS; Daniel Miller³, PhD; Jeannine Leverenz², MS, RN, CDCES; Korey Hood⁴,⁵, PhD; Ming Yeh Lee², MD, PhD; David M Maahs¹,⁴,⁵, MD, PhD; Priya Prahalad¹,⁴, MD, PhD

¹Department of Pediatrics, Division of Pediatric Endocrinology, Stanford University, Stanford, CA, United States
²Lucile Packard Children’s Hospital, Stanford University, Stanford, CA, United States
³Department of Management Science and Engineering, Stanford University School of Engineering, Stanford, CA, United States
⁴Stanford Diabetes Research Center, Stanford University, Stanford, CA, United States
⁵Department of Health Research and Policy, Stanford University, Stanford, CA, United States

Corresponding Author:
Priya Prahalad, MD, PhD
Department of Pediatrics, Division of Pediatric Endocrinology
Stanford University
300 Pasteur Dr.
Room G313
Stanford, CA, 94025
United States
Phone: 1 6507211811
Fax: 1 6507258375
Email: prahalad@stanford.edu

Abstract

Background: The use of continuous glucose monitors (CGMs) is recommended as the standard of care by the American Diabetes Association for individuals with type 1 diabetes (T1D). Few hardware-agnostic, open-source, whole-population tools are available to facilitate the use of CGM data by clinicians such as physicians and certified diabetes educators.

Objective: This study aimed to develop a tool that identifies patients appropriate for contact using an asynchronous message through electronic medical records while minimizing the number of patients reviewed by a certified diabetes educator or physician using the tool.

Methods: We used consensus guidelines to develop timely interventions for diabetes excellence (TIDE), an open-source hardware-agnostic tool to analyze CGM data to identify patients with deteriorating glucose control by generating generic flags (eg, mean glucose [MG] >170 mg/dL) and personalized flags (eg, MG increased by >10 mg/dL). In a prospective 7-week study in a pediatric T1D clinic, we measured the sensitivity of TIDE in identifying patients appropriate for contact and the number of patients reviewed. We simulated measures of the workload generated by TIDE, including the average number of time in range (TIR) flags per patient per review period, on a convenience sample of eight external data sets, 6 from clinical trials and 2 donated by research foundations.

Results: Over the 7 weeks of evaluation, the clinical population increased from 56 to 64 patients. The mean sensitivity was 99% (242/245; SD 2.5%), and the mean reduction in the number of patients reviewed was 42.6% (182/427; SD 10.9%). The 8 external data sets contained 1365 patients with 30,017 weeks of data collected by 7 types of CGMs. The rates of generic and personalized TIR flags per patient per review period were, respectively, 0.15 and 0.12 in the data set with the lowest average MG (141 mg/dL) and 0.95 and 0.22 in the data set with the highest average MG (207 mg/dL).

Conclusions: TIDE is an open-source hardware-agnostic tool for personalized analysis of CGM data at the clinical population scale. In a pediatric T1D clinic, TIDE identified 99% of patients appropriate for contact using an asynchronous message through electronic medical records while reducing the number of patients reviewed by certified diabetes care and education specialists by 43%. For each of the 8 external data sets, simulation of the use of TIDE produced fewer than 0.25 personalized TIR flags per
patient per review period. The use of TIDE to support telemedicine-based T1D care may facilitate sensitive and efficient guideline-based population health management.

(JMIR Diabetes 2022;7(2):e27284) doi: 10.2196/27284

KEYWORDS
telehealth; diabetes; population health; continuous glucose monitor; personalized medicine

Introduction

Background

For patients with type 1 diabetes (T1D) receiving insulin therapy, the American Diabetes Association (ADA) recommends the use of continuous glucose monitors (CGMs) as the standard of care along with quarterly clinic visits with hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) laboratory testing [1]. However, most people with T1D remain on self-monitored blood glucose because of patient, clinician, or insurance preference and do not meet the current HbA\textsubscript{1c} targets [2]. The long feedback cycle and the use of relatively little data when self-monitoring are barriers to timely detection and personalized response to deteriorating glucose control. An individual self-monitoring their glucose levels in line with the 2018 ADA standard of care recommendations generates glucose readings 6-10 times per day and receives feedback from their care team every 1–4 months based on a clinic visit or an HbA\textsubscript{1c} test [3]. In contrast, CGMs record glucose levels once every 5-15 minutes (96-288 times per day). The initiation and continued use of CGMs have increased and are associated with improved clinical outcomes and patient-reported quality of life measures [1-9]. In a US pediatric T1D registry, the use of CGMs increased from 4% in 2013 to 33% in 2017 [6].

Numerous commercial and open-source platforms provide individual-level visualizations and analyses of CGM data [10-13]. Recent studies by the Advanced Technologies and Treatments for Diabetes consensus on CGM and the ADA and European Association for the Study of Diabetes consensus on precision medicine in diabetes found that although the use of CGMs offers an opportunity to use high-frequency data to identify deteriorating glucose control and tailor personalized management strategies, no standardized, validated methods currently exist outside of automated insulin delivery systems [4,14]. Patient-level tools, such as manufacturer or data aggregator platforms, require physicians or certified diabetes care and education specialists (CDCESs) to examine the data of each individual to identify those people whose glucose management may need improvement. Population-level tools that analyze and present data for the entire population to facilitate prioritizing patients are less common. LibreView (Abbott Laboratories) enables whole-population data review but is proprietary, works only with Libre sensors, and provides access primarily to prespecified metrics [12]. A hardware-agnostic tool would be more appropriate to support care for a population in which patients use a variety of CGMs. An open-source tool would facilitate external evaluation and the development and comparison of alternative models. A tool that calculates personalized metrics for each patient based on their historical data would facilitate the tracking of temporal changes in glucose management. To the best of our knowledge, no validated, hardware-agnostic, open-source tool is available to facilitate the delivery of timely, population-level, personalized care through telehealth.

Numerous efforts have been made to improve T1D management using remote monitoring, the most successful of which relied on asynchronous messages sent to patients [15]. However, not all studies demonstrated significant and sustained improvement [15]. The implementation of clinical decision support (CDS) has faced a variety of challenges and has led to structured recommendations for their successful design and deployment [16]. Two of the primary areas of focus are that the CDS should improve, rather than disrupt, the appropriate workflows and that it should be designed with an iterative approach [16]. A recent multi-institution, cluster randomized clinical trial on the use of a CDS to improve the management of heart disease showed no significant improvements [17]. There was insufficient evaluation and redesign of the system based on feedback from its intended users [18]. The lack of iterative design and the resulting challenges to the workflow are common in the design of clinical software following the waterfall approach, a structured top-down approach in which the intent is to test and finalize the tool before deployment [19]. The agile approach, a more iterative approach based on rapid deployment and iterative redesign, is a popular alternative [19].

Objectives

We sought to design timely interventions for diabetes excellence (TIDE), a decision support tool to identify patients appropriate for asynchronous contact using a secure message through electronic medical record (EMR). To facilitate successful deployment and sustained use, we sought to fit into and improve current workflows by reducing the number of patients requiring review by a physician or the CDCES. We followed the agile approach to deploy the earliest viable version of TIDE in clinical practice and update it based on feedback from physicians and CDCESs.

Methods

Study Design

This study followed the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [20]. The first phase was hardware-agnostic algorithm design based on the data collected using a variety of CGM hardware. Data were obtained from a convenience sample of eight external data sets: 6 from clinical trials and 2 donated by research foundations [21-28]. The second phase was the design of TIDE, an interactive visual interface presenting CGM data for the entire clinical population, based on iterative feedback.
from physicians and CDCESs at an academic pediatric T1D clinic. One physician and CDCES used TIDE for 4 weeks and provided feedback based on which TIDE was comprehensively redesigned. Following this, they used the redesigned version for 5 weeks, during which minor improvements were made and errors were fixed. The third phase was a prospective evaluation over the course of 7 weeks of the sensitivity of TIDE for identifying patients who are appropriate for asynchronous contact using a secure message through the EMR and the difference in the number of patients requiring review with and without the use of TIDE. In the final phase, we simulated several measures of the workload generated by TIDE on the same convenience sample of 8 external data sets used to design the algorithms.

Setting
TIDE was developed at an academic pediatric T1D clinic caring for youth newly diagnosed with T1D who initiated CGM use within 1 month of onset and enrolled in a weekly remote monitoring program. All patients used Dexcom G6 (Dexcom) monitors, from which data were uploaded and made available to physicians and CDCESs through the Dexcom Clarity Clinic Portal [10]. Participants consented to participate in a longitudinal study evaluating the initiation of CGM early in the course of diabetes and the effects of weekly CGM data review as part of a larger ongoing study, for which the details of the consent process, eligibility criteria, screening, and enrollment process have been reported [8]. Each week, a CDCES used the Dexcom Clarity Clinic Portal to review each patient’s data and send an asynchronous message through the EMR to those patients who they determined required glucose management guidance. This study was approved by the Stanford University institutional review board. The leadership of the clinic, study authors PP and DMM, who serve on relevant national organizations governing diabetes technology, approved the use of TIDE in clinical care.

Generic and Personalized Metrics
A metric is generic if it is calculated the same way for each patient, for example, mean glucose (MG), and personalized if it is calculated for each patient based on their historical data; for example, the month-to-month change in MG. Consensus guidelines were used to generate a large set of generic CGM-based metrics from which clinicians could select the metrics to be tracked in TIDE [29]. The metrics included the number of days the CGM was active and collected more than a minimum percentage of valid readings (ACT), MG, percentage of time in range (TIR) defined as readings 70 to 180 mg/dL, percentage of time extremely hypoglycemic (eHyp) defined as readings <54 mg/dL, and percentage of time hypoglycemic (Hyp) defined as readings <70 mg/dL. Following the same consensus guidelines, each metric was calculated for each day for the entire day from 0:00 to 24:00, daytime 6:00 AM to midnight, and nighttime midnight to 6:00 AM. A complete list of generic metrics is provided in Table S1 in Multimedia Appendix 1. For each generic metric, a personalized metric was defined as the change from a baseline period to a review period (eg, the month-to-month change in the MG). The review period is the timeframe over which the metrics are calculated. It is defined relative to the day on which the data are being reviewed (eg, the last full week). The baseline period is the timeframe over which the baseline value for each personalized metric is calculated (eg, the last full month before the review period). For each metric, “a flag is triggered” when the metric exceeds a prespecified target value.

Algorithms to calculate generic and personalized metrics and to generate flags as a function of the review period, baseline period, and target value were developed and tested using data from 8 external data sets. The data sets were identified based on an internet search and the professional contacts of the authors: 6 previously published clinical trials and observational data donated by Tidepool and OpenAPS (Table S2 in Multimedia Appendix 1). Data use agreements were signed with Tidepool and OpenAPS, each stipulating that the data may be used for this research project and that those donating the data would not take part in the study design or the reporting of results and would be acknowledged in writing. The data sets had 168,723 patient days of included CGM readings collected with seven types of CGMs: Freestyle Navigator, Dexcom STS, Medtronic Paradigm or Guardian, iPro2, iPro2 Professional CGM, Freestyle Libre Pro Flash, and Dexcom G4. The algorithms with annotated codes and an overview of their design, TIDE, and synthetic CGM data for use with TIDE are available on GitHub [30,31].

Iterative Design of an Interactive Tool
The design of an initial version of TIDE was based on a convenience sample of informal interviews and observations used to establish the current state, achieve buy-in from stakeholders, and solicit suggestions for and perceived problems with the proposed workflow. The initial version of TIDE was designed to require a one-time setup followed by repeated use. During the one-time setup, based on their clinical practice and population, clinicians select the metrics to be displayed, the review period over which the generic metrics are calculated, and a baseline period based on which the personalized metrics are calculated. Two pediatric endocrinologists, study authors DMM and PP, and a CDCES, study author JL, identified the consensus glucose metrics currently being used in the clinic to evaluate patient glucose management: ACT was measured as the number of valid readings as a percentage of the maximum number of readings possible (the number of 5-minute intervals during the review period), MG, TIR, eHyp, and Hyp. The review period was set to 1 week ending on the last Sunday before the data review. The targets were initially set as follows: ACT >75%, TIR >70%, eHyp <1%, and Hyp <4%. No target was set for the MG. The valid wear threshold was used by the CDCES to determine whether to reach out to the patient to discuss their use of the CGM and to assist with challenges in obtaining additional sensors. When the valid wear threshold was not met, TIDE presented the metrics and flags as usual, and the CDCES used their judgment to determine whether patient data required further review.

During each of the first 4 weeks of the use of TIDE, the physician and study author PP logged into the Dexcom Clarity Clinic Portal and downloaded patient ID numbers, CGM readings, and CGM timestamps for all patients in the study. The physician used TIDE to identify patients with flags, reviewed
the output of TIDE for appropriateness and patient safety, and forwarded the list of patients flagged by TIDE to CDCES (study author JL). For each flagged patient, the CDCES opened Dexcom Clarity and reviewed detailed patient data, opened the EMR and sent a secure message to the patient, and, if a dose adjustment was made, the dose was updated in the patient’s chart. After these 4 weeks, PP and JL suggested changes that were incorporated into TIDE between weeks 4 and 5. During weeks 5 to 9, the tool was used by the study authors PP and JL to provide clinical care and identify minor adjustments required to improve usability or correct errors. Minor adjustments and corrections were made immediately, usually within 24 hours of identification.

Prospective Evaluation

The primary measures were the sensitivity of TIDE for identifying patients appropriate for asynchronous contact using a secure message through EMRs and the reduction in the number of patients reviewed by the CDCES with the use of TIDE. For 7 weeks, a CDCES reviewed the CGM data of each patient in the population and determined those appropriate for asynchronous contact. Sensitivity was defined as the number of appropriate patients flagged by TIDE divided by the number of appropriate patients. The reduction in the number of patients reviewed using TIDE was the number of patients not flagged by TIDE as a percentage of the total population.

The secondary outcomes calculated were as follows: the average amount of time required for per-patient review and contact as measured by the CDCES using TIDE, specificity (the number of patients not flagged by TIDE divided by the number of patients not appropriate for asynchronous contact), positive predictive power (the number of patients appropriately flagged by TIDE divided the number of patients flagged by TIDE), and negative predictive power (the number of patients not appropriate for asynchronous contact not flagged by TIDE divided by the number of patients not flagged by TIDE).

Validation on External Data Sets

The primary determinant of the workload associated with the use of TIDE is the number of patients for whom flags are generated in each review period, equivalent to the rate at which flags are generated per patient per review period. To evaluate the workload associated with the use of TIDE in other settings, we simulated the rate at which generic and personalized flags for MG and TIR were compared measured for each type of flag, each patient, and each dataset. For each data set, the rates at which flags were generated and the percentage of patients for whom at least one MG or TIR flag was generated every week were calculated. The rates of generic and personalized flags for MG and TIR were compared using a 2-tailed paired t test. The number of patients for whom an MG or TIR flag was generated every week was recorded.

Results

Iterative Design of an Interactive Tool

The first version of TIDE displayed the metrics that triggered a flag in red and those that did not trigger a flag in green (Figure 1). After the feedback from the first 4 weeks of its use, the primary changes to the interface were as follows: the columns displaying MG, the number of readings, and the number of 5-minute intervals during the review period were removed to minimize the number of patients who received flags that did not require dose adjustments while not missing those who required dose adjustments; the criterion for TIR target was changed from 70% to 60%; a personalized metric was added to compare each patient’s TIR in the previous week to their TIR in the previous 4 weeks with a target of an increase in TIR or a drop in TIR of no more than 10% points; the color-coding was revised so only metrics that triggered a flag were highlighted; the wording of the display names of the metrics was changed to be more interpretable; a feature was added to allow the person using the tool to specify whether to use data from the most recent 7 days or from the default review period of 7 days ending on the previous Sunday; the patient data, previously presented on a single tab, were split tab into four tabs that displayed all patients, patients with alerts, patients with no data, and patients with data but no alerts; and the visual presentation was made more compact to display more patients per page (Figure 1).

The primary change in the workflow was that the step of downloading each patient’s data from Dexcom Clarity was replaced with a Python script that downloaded all patient data (Figure 2). The participation of the physician in the review process was no longer necessary as part of the workflow but was continued to ensure patient safety and quality of care. The ultimate intended workflow, initiated months after the completion of this study, is for the CDCES to use TIDE without the participation of physicians.
Prospective Evaluation

Over the last 7 weeks of the study, the number of patients increased from 56 to 64, totaling 427 patient weeks. The sensitivity of TIDE for identifying patients appropriate for contact using an asynchronous message through the EMR was 94% in the first week, 96% in the last week, and 100% in all other weeks (mean 99%, SD 2.5%; Table 1). The average
reduction in the number of patients reviewed by the CDCES was 42.8% (182/427; SD 10.9%), that is, the fraction of patients not flagged for review by TIDE (Table 1).

For patients identified by TIDE as requiring review, the mean duration of the data review process averaged 4.5 minutes per patient (1.5 minutes to access the data in Dexcom Clarity and review it for patterns, 1 minute to log into the patient’s record and document changes, 2 minutes to send the patient a message using a secure EMR-based messaging platform). The weekly specificity, positive predictive power, and negative predictive power of TIDE are shown in Table 1.

Table 1. Outcomes including sensitivity identifying patients appropriate for asynchronous contact through the medical record and the reduction in the number of patients reviewed.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients in study, N</td>
<td>363</td>
<td>373</td>
<td>393</td>
<td>393</td>
<td>292</td>
<td>292</td>
<td>292</td>
</tr>
<tr>
<td>Patients flagged by TIDE, n</td>
<td>56</td>
<td>58</td>
<td>61</td>
<td>63</td>
<td>61</td>
<td>58</td>
<td>56</td>
</tr>
<tr>
<td>Patients reviewed, %</td>
<td>63</td>
<td>62</td>
<td>39</td>
<td>37</td>
<td>39</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td>Reduction in patients reviewed, %</td>
<td>34</td>
<td>34</td>
<td>30</td>
<td>27</td>
<td>31</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>True positive flags, %</td>
<td>29</td>
<td>21</td>
<td>31</td>
<td>27</td>
<td>30</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>True negative flags, %</td>
<td>23</td>
<td>17</td>
<td>19</td>
<td>15</td>
<td>9</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>False positive flags, %</td>
<td>23</td>
<td>33</td>
<td>36</td>
<td>40</td>
<td>39</td>
<td>33</td>
<td>31</td>
</tr>
<tr>
<td>False negative flags, %</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Insufficient CGM data, %</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td>94</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Specificity, %</td>
<td>50</td>
<td>34</td>
<td>34</td>
<td>26</td>
<td>19</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>Positive predictive value, %</td>
<td>55</td>
<td>39</td>
<td>46</td>
<td>40</td>
<td>43</td>
<td>51</td>
<td>52</td>
</tr>
<tr>
<td>Negative predictive value, %</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>92</td>
</tr>
</tbody>
</table>

aTIDE: timely interventions for diabetes excellence.

bPrimary objective (italicized).

cFlagged by TIDE and appropriate for asynchronous contact.

dNot flagged by TIDE and not appropriate for asynchronous contact.

eFlagged by TIDE and not appropriate for asynchronous contact.

fNot flagged by TIDE and appropriate for asynchronous contact.

gCGM: continuous glucose monitor.

Validation on External Data Sets

There were 1424 patients with at least 1 day of CGM data that met the inclusion criteria in the 8 external data sets. There were 168,723 patient days with CGM readings across 30,076 weeks. The patient with the most included days had 1028 days over 154 weeks, whereas the patient with the fewest days had 1 day. The mean weekly ACT was 5.1 (IQR 4.0-6.33), MG was 170.7 (IQR 148.8-189.2), mean percentage of TIR 70 to 180 mg/dL was 56.6% (IQR 45.1%-68.4%), eHyp <54 mg/dL was 1.9% (IQR 0.25%-3.23%), and Hyp <70 mg/dL was 3.3% (IQR 1.50%-4.60%). Across data sets, the minimum and maximum number of patients included were 12 and 450, respectively; the number of CGM days per patient was 9.8 and 256.8, respectively; MG was 141.3 (SD 21.3) and 207.0 (SD 35.3), respectively; mean percentage TIR was 38.2% (SD 14.2%) and 74.2% (SD 13.3%), respectively; mean percentage TIR was 38.2% (SD 14.2%) and 74.2% (SD 13.3%), respectively; and eHyp was 2.1% (SD 2.3%) and 4.6% (SD 3.8%), respectively; and Hyp was 2.1% (SD 2.3%) and 4.6% (SD 3.8%), respectively; and Hyp was 2.1% (SD 2.3%) and 4.6% (SD 3.8%), respectively (Table 2). Data sets were numbered by increasing mean MG. The analysis of personalized MG and TIR flags included 1100 patients with at least 4 weeks of data. The median frequency was significantly higher for generic TIR flags than for personalized TIR flags, 0.47 (IQR 0.12-0.83) versus 0.19 (IQR 0.12-0.26) flags per patient per week ($P<.001$), respectively, as was the SD of the frequency of flags (0.36 vs 0.11, respectively; $P<.001$; Figure 3). The median frequency of flags was significantly higher for generic MG than for personalized MG, 0.31 (IQR 0.05-0.71) versus 0.30 (IQR 0.22-0.37) flags per patient per week ($P<.001$), as was the SD of the frequency of flags 0.35 versus 0.12 ($P<.001$; Figure 3). In the two data sets with the highest mean MG, data sets 7 and 8, respectively, 49.7% (97/212) and 81% (26/32) of patients had a generic MG or TIR flag every week. Across all data sets, 15.64% (172/1100) of patients had a generic MG or TIR flag every week, and 0% (1/1100) of patients had a personalized MG or TIR flag every week (Table 3).
Table 2. Continuous glucose monitor (CGM) data in external data sets.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Patients, n</th>
<th>CGM days, n (number per patient)</th>
<th>CGM weeks, n (number per patient), n</th>
<th>Days active, mean (SD)</th>
<th>Glucose (mg/dL), mean (SD)</th>
<th>Time in range(^a) (%), mean (SD)</th>
<th>Percentage of time hypoglycemic(^b), mean (SD)</th>
<th>Percentage of time extremely hypoglycemic(^c), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85</td>
<td>21,236 (249.8) 3372 (39.7)</td>
<td>6.3</td>
<td>141.3 (21.3)</td>
<td>74.2 (13.3)</td>
<td>3.1 (2.4)</td>
<td>2 (3.1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>315 (26.2) 63 (5.2)</td>
<td>5</td>
<td>152.9 (23.3)</td>
<td>66.4 (13.2)</td>
<td>4 (3)</td>
<td>1.7 (2.4)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>30,815 (256.8) 4546 (37.9)</td>
<td>6.8</td>
<td>158.3 (26.7)</td>
<td>65.9 (16)</td>
<td>2 (1.3)</td>
<td>0.5 (0.9)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>225</td>
<td>48,805 (216.9) 7670 (34.1)</td>
<td>6.4</td>
<td>162.3 (24.1)</td>
<td>62.1 (14.1)</td>
<td>3 (2.6)</td>
<td>0.9 (1.4)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>450</td>
<td>54,535 (121.2) 11,936 (26.5)</td>
<td>4.6</td>
<td>162.6 (28)</td>
<td>62.4 (16.3)</td>
<td>2.9 (3.1)</td>
<td>1 (2.3)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>180</td>
<td>1682 (9.3) 378 (2.1)</td>
<td>4.4</td>
<td>176.6 (33.3)</td>
<td>50.4 (15.6)</td>
<td>3.6 (3.2)</td>
<td>3.4 (4.3)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>219</td>
<td>10,025 (45.8) 1727 (7.9)</td>
<td>5.8</td>
<td>184.9 (44.4)</td>
<td>45.1 (16.6)</td>
<td>4.6 (3.8)</td>
<td>3.7 (4.9)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>133</td>
<td>1310 (9.8) 384 (2.9)</td>
<td>3.4</td>
<td>207 (35.3)</td>
<td>38.2 (14.2)</td>
<td>2.9 (3.7)</td>
<td>1.6 (3.6)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Percentage of readings that were 70 to 180 mg/dL.

\(^b\)Percentage of readings <70 mg/dL.

\(^c\)Percentage of readings <54 mg/dL.

Figure 3. Frequency of generic and personalized flags in external cohorts.
rather than spending time scheduling an appointment or trying to contact the patient or family. The primary challenges of the agile approach and asynchronous messaging are that additional resources may be required to ensure that patient care is not adversely affected during the deployment of an early stage tool or because of patients ignoring messages. In this study, during the initial testing period, a physician reviewed the output of TIDE to ensure that the quality of care was not compromised, and the CDCES tracked whether patients read their messages and followed up accordingly.

The use of generic and personalized flags has complementary benefits. For patients with average glucose management, the generic flags provide a standardized approach to care based on the most recent consensus guidelines. For patients with very well or very poorly managed glucose levels, personalized flags based on patient progress may be more informative. If a person with MG 208 (the mean in one of the external data sets) consistently reduced their glucose by 5 mg/dL per review period, a generic metric may trigger a flag for numerous consecutive review periods, whereas a personalized metric would indicate improvement. During the first major revision of TIDE, a personalized metric was added to track the TIR to help CDCESs identify changes in patient management that did not cross the threshold of a personalized metric. With any flags, particularly personalized metric was added to track the TIR to help CDCESs identify changes in patient management that did not cross the threshold of a personalized metric. With any flags, particularly

### Discussion

#### Principal Findings

We designed TIDE as an open-source hardware-agnostic tool for the personalized analysis of CGM data at the clinic population scale. In a pediatric T1D clinic, TIDE identified 99% of patients appropriate for contact using an asynchronous message through the EMR while reducing the number of patients reviewed by certified diabetes educators by 43%. For each of the 8 external data sets, simulation of the use of TIDE produced fewer than 0.25 personalized TIR flags per patient per review period. TIDE and the underlying algorithms are free open-source software available on GitHub [30,31]. Upon request, we will help clinicians customize the tool to their setting and deploy it in practice.

TIDE was developed with an iterative agile approach to support asynchronous contact with patients, the form of contact found to be most effective by a systematic review of T1D telemetry systems [15]. An initial version of TIDE was produced quickly, and few significant design decisions were made before physicians and CDCESs used TIDE to provide feedback. Physicians and CDCESs identified their preferences for a rule-based approach over less-interpretable approaches such as machine learning, time-series analysis, or alternative statistical smoothing techniques. On the basis of physician and CDCES feedback, TIDE was designed to identify how a patient’s glucose management differs from validated recommendations and to produce interpretable flags to facilitate recommendations. As TIDE uses consensus guidelines, it may be more broadly applicable than a model trained on a small or nonrepresentative subset of the population would be. The use of TIDE to identify patients for asynchronous messaging fits well with the CDCES workflow. The CDCES could send a message to each patient identified by TIDE and move immediately to the next patient, rather than spending time scheduling an appointment or trying

### Table 3. Frequency of generic and personalized flags based on continuous glucose monitor data in external data sets.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Patients, n</th>
<th>Patients with at least 4 weeks of data, n (%)</th>
<th>Glucose flags, n (flags per patient per week)</th>
<th>Time in range flags&lt;sup&gt;a&lt;/sup&gt;, n (flags per patient per week)</th>
<th>Patients with mean glucose or time in range flag every week, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Generic&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Personalized&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Generic&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>1</td>
<td>85</td>
<td>78 (92)</td>
<td>329 (0.1)</td>
<td>725 (0.22)</td>
<td>493 (0.15)</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>10 (83)</td>
<td>13 (0.22)</td>
<td>14 (0.24)</td>
<td>19 (0.33)</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>120 (100)</td>
<td>1417 (0.31)</td>
<td>1128 (0.25)</td>
<td>1535 (0.34)</td>
</tr>
<tr>
<td>4</td>
<td>225</td>
<td>225 (100)</td>
<td>2595 (0.34)</td>
<td>2002 (0.26)</td>
<td>3204 (0.42)</td>
</tr>
<tr>
<td>5</td>
<td>450</td>
<td>436 (97)</td>
<td>4137 (0.35)</td>
<td>3498 (0.29)</td>
<td>4895 (0.41)</td>
</tr>
<tr>
<td>6</td>
<td>180</td>
<td>4 (2)</td>
<td>6 (0.38)</td>
<td>4 (0.25)</td>
<td>9 (0.56)</td>
</tr>
<tr>
<td>7</td>
<td>219</td>
<td>195 (89)</td>
<td>967 (0.58)</td>
<td>525 (0.31)</td>
<td>1360 (0.81)</td>
</tr>
<tr>
<td>8</td>
<td>133</td>
<td>32 (24)</td>
<td>116 (0.84)</td>
<td>47 (0.34)</td>
<td>131 (0.95)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Personalized mean glucose (MG) flag triggered when MG > MG + 10 mg/dL in baseline period.

<sup>b</sup>Generic mean glucose flag triggered when mean bigeneric glucose > 170 mg/dL.

<sup>c</sup>Generic time in range flag triggered when the percentage of readings of 70-180 mg/dL was <60%.

<sup>d</sup>Personalized time in range (TIR) flag triggered when TIR < TIR – 10% points in baseline period.

---

https://diabetes.jmir.org/2022/2/e27284
of a relatively simple tool with metrics and targets based on the consensus guidelines may be useful for nonspecialist clinicians to inform the care of such patients.

**Strengths and Limitations**

A strength of this study is the evaluation across 8 external datasets of the potential applicability of TIDE. Most T1D clinics in the United States see patients 3 to 12 times per year primarily in person [3]. In this study, reviewing a patient’s data with TIDE and sending the patient a message required an average of 4.5 minutes. In a clinic in which patients use CGM and a 15-minute-long in-person visits, eliminating an average of 1 in-person visit per patient per year would provide sufficient time for an average of 3 message-based contacts. Using TIDE to support such a workflow requires that TIDE flags an average of 0.25 patients per month for review, equivalent to 3 reviews per patient per year. In a simulation of the use of TIDE for populations with differing average levels of glucose management, the personalized TIR metric flagged no more than 0.22 patients per review period, even for the population with an average MG level of 207 mg/dL. As TIDE flags patients based on deteriorating glucose control, patient contacts would be targeted to address the need rather than per a fixed schedule. Each patient would be more likely to receive care when their control deteriorates. On an average, patients with worse control may receive more contacts than those with better control. Such deployments of TIDE are ongoing with two partner clinics, one in the United States and another in Australia, each caring for ≥1000 patients with T1D and using CGMs.

This study has several limitations. The workflow presented requires downloading data and toggling between the tool and Dexcom Clarity and is not integrated with the EMR. Integration of CGM data with the EMR will facilitate integration with current telehealth workflows and allow the tool to incorporate data on the timing of each patient’s previous and upcoming visits into the recommendations. The specificity of this tool was significantly lower than its sensitivity. However, specificity is less relevant than the direct measure of the primary outcome and time savings associated with the use of TIDE. Specificity may be improved by using the data on which patients did and did not require a review to tune the algorithm for which patients should be flagged. Improvements are ongoing to streamline, standardize, and scale the data review; to improve the sensitivity and specificity of the tool; and to incorporate the tool into the EMR [35]. There was a 43% reduction in the number of patients reviewed each week; however, reviewing patients weekly is not currently the standard of care and may represent increased time spent for most diabetes clinicians. The reduction in the number of patients requiring review in a different setting may change with the cadence of the review and the criteria for review. This study was conducted as a novel proof-of-concept intervention to create new knowledge and generate data on the performance of an automated tool. The metrics and thresholds were not derived from a systematic hypothesis-based approach or a survey of patients, families, and clinicians. The thresholds used to create the tool were based on consensus guidelines with the input of a group of experts in diabetes technology, and the revised version of the tool used feedback from 4 weeks of use in clinical care.

**Future Studies**

Subsequent efforts to deploy TIDE in other settings may benefit from a formal quality improvement framework with an established aim, predefined targets for metrics, implementation criteria, and well-defined iteration cycles. Subsequent research is ongoing to explicitly incorporate measurements of the additional time necessary for clinical decision-making based on the review of the data, examine the operational requirements to expand the number of individuals monitored with the help of this tool, and identify if the use of such a tool may allow a clinic with fixed resources to provide care for more patients through a more efficient use of clinician time [35].

**Conclusions**

We developed and deployed TIDE, a tool that uses metrics based on consensus guidelines, to identify 99% of patients appropriate for contact using an asynchronous message while reducing the number of patients requiring review by a physician or certified diabetes educator by 43%. Further investigation is necessary to understand the potential of automated analyses of CGM data to support broader access to personalized and timely glucose management.

**Acknowledgments**

The authors thank the staff of the type 1 diabetes (T1D) clinic in which the study was conducted. The authors thank Tidepool (Howard Look and Brandon Arbiter) and OpenAPS (Dana Lewis) for sharing data for these analyses. For the data used from the T1D Exchange, the analyses, content, and conclusions presented herein are the sole responsibility of the authors and have not been reviewed or approved by the T1D Exchange.

We thank the 4T Study team including Ananta Addala, Dessi Zaharieva, Nora Arrizon-Ruiz, Franziska Bishop, Annette Chmielewski, Barry Conrad, Ana Cortes, Manisha Desai, Victoria Ding, Rebecca Gardner, Simrat Ghuman, Carolyn Herrera, KH, Julie Hooper, Brianna Leverenz, JL, Erica Pang, Natalie Pageler, Piper Sagan, Julie Senaldi, Anjoli Martinez-Singh, and Michelle Wiedemann.

We thank the numerous engineering students and physicians that contributed to the technical design of TIDE, the design of the operational model, and its iterative improvement: Johannes O Ferstad, Jacqueline J Vallon, Daniel Jun, Angela Gu, Anastasiya Vitko, Dianelys P Morales, MYL, Christos Vasilakis, Andrew Shin, and Esli Osmanliu.

This study was supported by R18DK122422, and PP was supported by P30DK116074.
Conflicts of Interest

DM has research support from the National Institutes of Health, JDRF, National Science Foundation, and the Helmsley Charitable Trust, and his institution has research support from Medtronic, Dexcom, Insulet, Big Foot Biomedical, Tandem, and Roche. DM consulted Abbott, Helmsley Charitable Trust, Sanofi, Novo Nordisk, Eli Lilly, and Insulet. KH has research support from Dexcom, Inc for investigator-initiated research and consultant fees from the Lilly Innovation Center; Lifesystems Diabetes Institute; Insulet, Inc; and Roche Diagnostics. AW receives compensation from HealthPals, Inc. in the form of salary. AW was supported by the US Department of Defense through a National Defense Science and Engineering Graduate (NDSEG) Fellowship.

Multimedia Appendix 1

Full list of generic metrics, sources of data for retrospective analysis, and timely interventions for diabetes excellence screenshots. [PDF File (Adobe PDF), 222 KB-Multimedia Appendix 1]

References

10. Dexcom Clarity. URL: https://clarity.dexcom.com [accessed 2021-05-01]


23. Tidepool. URL: https://tidepool.org [accessed 2022-03-16]


31. SURF-TIDE. GitHub. 2021 Sep 15. URL: https://github.com/joshuagrossman/a1c [accessed 2022-03-16]


Abbreviations

ACT: number of days continuous glucose monitor active
ADA: American Diabetes Association
CDCES: certified diabetes care and education specialist
CDS: clinical decision support
CGM: continuous glucose monitor
eHyp: extreme hypoglycemic
EMR: electronic medical record
Hyp: hypoglycemic
MG: mean glucose
T1D: type 1 diabetes
TIDE: timely interventions for diabetes excellence
TIR: time in range

©David Scheinker, Angela Gu, Joshua Grossman, Andrew Ward, Oseas Ayerdi, Daniel Miller, Jeannine Leverenz, Korey Hood, Ming Yeh Lee, David M Maahs, Priya Prahalad. Originally published in JMIR Diabetes (https://diabetes.jmir.org), 06.06.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Diabetes, is properly cited. The complete bibliographic information, a link to the original publication on https://diabetes.jmir.org/, as well as this copyright and license information must be included.