

ORIGINAL ARTICLE

Assessing Sensor Accuracy for Non-Adjunct Use of Continuous Glucose Monitoring

Boris P. Kovatchev, PhD,¹ Stephen D. Patek, PhD,² Edward Andrew Ortiz, MD,¹ and Marc D. Breton, PhD¹

Abstract

Background: The level of continuous glucose monitoring (CGM) accuracy needed for insulin dosing using sensor values (i.e., the level of accuracy permitting non-adjunct CGM use) is a topic of ongoing debate. Assessment of this level in clinical experiments is virtually impossible because the magnitude of CGM errors cannot be manipulated and related prospectively to clinical outcomes.

Materials and Methods: A combination of archival data (parallel CGM, insulin pump, self-monitoring of blood glucose [SMBG] records, and meals for 56 pump users with type 1 diabetes) and in silico experiments was used to “replay” real-life treatment scenarios and relate sensor error to glycemic outcomes. Nominal blood glucose (BG) traces were extracted using a mathematical model, yielding 2,082 BG segments each initiated by insulin bolus and confirmed by SMBG. These segments were replayed at seven sensor accuracy levels (mean absolute relative differences [MARDs] of 3–22%) testing six scenarios: insulin dosing using sensor values, threshold, and predictive alarms, each without or with considering CGM trend arrows.

Results: In all six scenarios, the occurrence of hypoglycemia (frequency of BG levels ≤ 50 mg/dL and BG levels ≤ 39 mg/dL) increased with sensor error, displaying an abrupt slope change at MARD = 10%. Similarly, hyperglycemia (frequency of BG levels ≥ 250 mg/dL and BG levels ≥ 400 mg/dL) increased and displayed an abrupt slope change at MARD = 10%. When added to insulin dosing decisions, information from CGM trend arrows, threshold, and predictive alarms resulted in improvement in average glycemia by 1.86, 8.17, and 8.88 mg/dL, respectively.

Conclusions: Using CGM for insulin dosing decisions is feasible below a certain level of sensor error, estimated in silico at MARD = 10%. In our experiments, further accuracy improvement did not contribute substantively to better glycemic outcomes.

Introduction

CONTINUOUS GLUCOSE MONITORING (CGM) is a powerful tool assisting the optimization of glycemic control in diabetes. Since the advent of CGM technology,^{1–3} significant progress has been made toward versatile and reliable devices that not only approximate the course of blood glucose (BG) fluctuations day and night, but also provide feedback such as alarms when preset low or high thresholds are reached. Several studies have documented the benefits of CGM^{4–7} and charted guidelines for its clinical use^{8,9} and for its future as a base for closed-loop control.^{10,11}

Physiology and CGM errors

It is important to note that subcutaneous CGM devices measure glucose concentration in a compartment different than blood—the interstitium—and then deduce BG concentration from interstitial glucose (IG) readings. Presumably, IG fluctuations are related to BG via a diffusion process, which results in a well-defined codependence allowing BG changes to be deduced from IG dynamics.^{12–14} To account for the gradient between BG and IG, CGM glucose is calibrated using capillary glucose measurements to match CGM glucose and BG levels.

¹University of Virginia Center for Diabetes Technology, Charlottesville, Virginia.

²Department of Systems and Information Engineering, University of Virginia, Charlottesville, Virginia.

Successful calibration would adjust the amplitude of IG fluctuations with respect to BG but would not completely eliminate the time lag due to BG-to-IG transport and instrument delay. Because the time lag can greatly influence the accuracy of CGM, several studies were dedicated to its investigation, yielding various results.^{15–17} For example, it was hypothesized that if a fall in glucose level is due to peripheral glucose consumption, the physiological time lag would be negative (i.e., fall in IG would precede fall in BG^{12,17}). In most studies IG lagged behind BG by 4–10 min regardless of the direction of BG change. The push–pull phenomenon offered reconciliation of these results,¹⁸ and a recent precise measurement settled the time lag in fasting overnight state to 5–6 min.¹⁹

In addition, errors from calibration, transient loss of sensitivity, and random noise confound CGM data.²⁰ Nevertheless, the accuracy of CGM is increasing and may be approaching a physiological limit for subcutaneous glucose monitoring.^{21–27}

CGM data and information

CGM generates data streams that are both voluminous and complex. From an analytical point of view, these data are *time series*—sequences of BG readings that are equally spaced in time (e.g., every 5 min). This data representation results in both challenges to the assessment of sensor accuracy and advantages in terms of information processing. For example, the use of metrics such as mean absolute relative difference (MARD) or correlation between CGM and reference BG data as a sole source for accuracy assessment *undervalues substantially the information carried by CGM about the glucose fluctuation process*.

A broader concept of CGM accuracy based not only on the deviation of CGM readings from reference BG, but also on the device's ability to follow glucose trends, was presented in 2004 with the introduction of the Continuous Glucose Error-Grid Analysis (CG-EGA)²⁸ and was then adopted in a performance metrics guideline by the Clinical and Laboratory Standards Institute.²⁹ It was proposed that the additional knowledge of BG rate and direction of change offered by CGM, albeit imperfect, adds valuable information to the recognition of the real glycemic state of a person. The CG-EGA was used to compare the clinical utility of the information provided by CGM to the information provided by episodic self-monitoring (self-monitoring of BG [SMBG]), concluding that “While SMBG produces more accurate instantaneous glucose values than CGM...the additional information about the direction and rate of glucose change increases the ability to make correct clinical decisions when compared to episodic SMBG tests.”³⁰ In other words, judging CGM accuracy with the same metrics that are used for accuracy assessment of episodic SMBG underestimates the information carried by CGM data.

This topic was extensively discussed in two review articles: one focusing on the statistical aspects of CGM data analysis,³¹ and the other on a comprehensive engineering account of the signal and models used by diabetes technology, including closed-loop control.³²

Utility of CGM information

Outcome studies demonstrate that adding CGM information results in improved glycemic control with sensor-

augmented insulin pump therapies.^{33,34} In addition to presenting frequent BG data, CGM devices typically display directional trends and generate alarms and warning messages for hypo- and hyperglycemia. These features have rapidly evolved from a concept³⁵ to implementation in CGM systems, and the next logical step was successfully undertaken—automated mitigation of hypoglycemia via shutoff of the insulin pump when a glucose threshold was reached.^{36–38} Predictive algorithms bring the CGM information processing to a higher level,³⁹ and closed-loop control, known as the artificial pancreas, offers the ultimate retrieval of CGM information via complex algorithms, typically based on elaborate models of the human glucose metabolism. Between 2008 and 2011, promising results from inpatient closed-loop control studies were reported by several groups.^{40–45} The results from these investigations were summarized in a 2011 review of the artificial pancreas field,⁴⁶ pointing out the superiority of closed-loop control over insulin pump therapy in terms of (1) increased time within target glucose range, (2) reduced incidence of hypoglycemia, and (3) better overnight control. Subsequent studies confirmed these findings in outpatient setting,⁴⁷ at diabetes camps for children,⁴⁸ and at patients' homes.⁴⁹ The superior glucose control achieved by these trials suggests that when CGM information is processed appropriately by advanced algorithms, the clinical outcomes are better, even than those achieved by state-of-the-art sensor-augmented insulin pump therapy.^{50,51}

Thus, CGM data contain valuable information reaching beyond the episodic BG determinations provided by SMBG. At a certain level of CGM accuracy, this information should permit effective “non-adjunct” use of CGM as a replacement for SMBG. In this article we use real data and extensive computer simulation to gauge the level of CGM accuracy that would allow non-adjunct CGM use in three scenarios: (1) direct point-of-care (POC) insulin dosing decisions made using CGM values; (2) hypo- and hyperglycemia detection and threshold alarms; and (3) hypo- and hyperglycemia prediction and predictive alarms. Each of these scenarios is considered without or with use of the trend arrow display typically available on a CGM device screen. We omit intentionally the case of closed-loop control because its effectiveness depends heavily on the performance of the control algorithm.

Materials and Methods

The basic idea of our method is to use real-life “true” BG traces of sufficient duration for patients with diabetes and then replay in silico “what-if” treatment scenarios gradually increasing the sensor error. This approach allows the clinical outcomes observed in real life to be augmented by events due to sensor errors, thereby linking increasing sensor inaccuracy to deterioration in glycemic control. The elements of this approach are as follows:

Database

This study used archived de-identified data collected during the project funded by National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases grant RO1 DK 085623 (see clinicaltrials.gov for clinical trial registration number NCT01434030). Sixty insulin pump users with type 1 diabetes were recruited and

were asked to wear a CGM device for a month, simultaneously recording SMBG, CGM, and insulin pump data, as well as information about meals and physical activity. Fifty-six participants completed the data collection and contributed to the database. The demographic characteristics of these subjects were as follows: 21–65 years of age, with a mean (SD) age of 41 (12.2) years; duration of type 1 diabetes of at least 2 years, with a mean duration of 24.1 (11.0) years; use of an insulin pump for at least 6 months, with a mean interval of 10 (5.8) years; and active use of a bolus calculator function. Mean (SD) hemoglobin A1c level was 7.7% (1.2%), 59% were female, the majority (95%) were white, and 50% were employed in professional occupations. The database building protocol was approved by the local Institutional Review Board; study details were previously published.⁵²

Nominal BG traces

To retrieve “true” BG fluctuation from SMBG, CGM, insulin pump, and meal data, we used an established model of glucose–insulin metabolism, which allowed the extraction of nominal BG traces for each person. Although a detailed description of this model is beyond the scope of this article, the procedure worked as follows:

1. Each subject’s CGM trace was smoothed and retrofitted to pass through the recorded SMBG values.⁵³ The retrofit procedure consists of two sequential steps: (1) smoothing of the raw sensor data and (2) adjustment of the amplitude of the smoothed sensor data to map the CGM trace to SMBG values. In situations where multiple fingerstick values were recorded within 10 min of each other, the sample that was numerically closest to the prevailing CGM value was used in the retrofit procedure.
2. Each subject’s actual basal rates and delivered insulin boluses were retrieved from the insulin pump.
3. A “net effect” signal was computed for each CGM-day for each subject, representing the best least squares estimate of the carbohydrate arrival process that explains the relationship between retrofitted CGM and the record of insulin delivery. Specifically, the net effect signals were computed retrospectively (on a daily basis) to be the estimated carbohydrate arrival process that minimizes an objective function that penalizes both (1) sum-of-squares differences between the model-predicted interstitial BG and retrofitted CGM and (2) sum-of-squared net effect values, where the latter is a “regularization” term that has the effect of smoothing the estimated carbohydrate trace. The resulting estimate is referred to as “net effect” because it is the only mechanism by which BG variability can be explained—meals are certainly reflected in the result, but other artifacts also appear: persistent insulin resistance shows up as a positive net effect, and sudden increases in insulin sensitivity (e.g., due to physical activity) can show up as negative values of net effect. The mathematical model used to compute the net effect traces is a subcutaneous oral glucose minimal model, which has been used in previous studies.⁵⁴

By construction, the recovered net effect signal is such that when it is “fed” to the subcutaneous oral glucose minimal

model along with the corresponding history of actual insulin delivery the model recovers the original CGM trace. The ability to “recover” the CGM trace with the actual record of insulin delivery becomes an ability to *simulate* the effect of modified insulin delivery, a technique referred to here as *net effect simulation*. A Net Effect Simulator was constructed from the patient database described above, where net effect signals were extracted from 2,082 nominal BG traces including bolus episodes. The lengths of the traces were determined to correspond roughly to the interval between calibrations for current CGM systems, with average length of 11 h. Each bolus episode included in the Net Effect Simulator represents an actual bolus event captured in the field data, beginning with a bolus that was accompanied by a corresponding fingerstick SMBG value and including all boluses delivered in the next 10–12 h (with or without accompanying SMBG readings). There is no overlap of the episodes included in the database. Bolus episodes were assembled sequentially over time, skipping CGM data gaps longer than 3 h and bridging smaller gaps with cubic spline interpolation.

Sensor error

A defining characteristic of the sequence of sensor errors in any CGM trace is that sequential errors are *not independent* (i.e., cannot be represented by simple random noise²⁰). In other words, when a sensor is incorrect, the error would persist for a while in the same direction. Thus, sensor errors are characterized by both their magnitude and by their autocorrelation. The interpretation of error magnitude is straightforward, and its numerical assessment is done by accepted metrics such as MARD. The interpretation of autocorrelation is more complex: roughly, higher autocorrelation would result in longer stretches of sensor errors in the same direction, observed for example during transient loss of sensitivity.⁵⁵ To simulate realistic CGM errors, we used 280 actual CGM traces (12 h long) compared with frequent YSI glucose analyzer measurement (every 15 min). Each trace produces a “signature” in term of fluctuation of the error in time (time-dependent bias, drift, delay, low-frequency fluctuations, and high-frequency variability). For each simulation we chose a specific CGM error signature from the pool of 280 and applied it to the desired plasma glucose trace multiplied by a predetermined factor; this way we could “dial” up or down the magnitude of sensor error while preserving all other observed noise characteristics. As a result, the temporal structure of sensor errors was preserved, allowing for clean manipulation of the error magnitude.

Replay of observed treatment scenarios

In our data, each nominal BG trace encompassed meals and included insulin dosing done using SMBG values. We replayed the patients’ treatment actions in silico, with insulin dosing done using CGM readings according to three treatment modalities, leading to a change in insulin dosing commensurate with the CGM errors:

1. POC use of CGM in which insulin dosing decisions at the same decision points as in the nominal traces are made on CGM values using the carbohydrate ratios and correction factors each specific patient used at that time.

2. Threshold alarms: mitigation of hypoglycemia and hyperglycemia by detection based on threshold “alarms” issued when the CGM values reach 70 mg/dL and 180 mg/dL, respectively.
3. Predictive alarms: Alarms for hypoglycemia are issued when a linear prediction of CGM glucose 20 min ahead reached a 70 mg/dL threshold, and alarms for hyperglycemia are issued when a linear prediction of CGM glucose 30 min ahead reached a 180 mg/dL threshold.

Each of these treatment modalities was tested at seven levels of sensor error ranging from MARD=3% to MARD=22%, yielding about 15,000 simulated CGM traces per treatment modality. The response to the hypoglycemia threshold or predictive alarms was a simulated treatment with 15 g of fast-acting carbohydrates every 15 min until the alarm condition resolved. Mandatory treatment with 30 g of carbohydrate was given for true BG levels ≤ 39 mg/dL regardless of sensor alarm. The response to hyperglycemia threshold or predictive alarms was an insulin correction that took into account each subject’s individual correction factor at that time of day and insulin injection history to restore euglycemia (120 mg/dL), giving 50% of such a correction every hour until the alarm condition resolved. In addition, each treatment modality was simulated twice, without and with taking into account the trend arrow presented by the CGM. Insulin doses were increased for arrows going up and decreased for arrows going down as clinically indicated.⁹ Finally, each period was simulated five times to present several random sensor profiles to each observed clinical situation, leading to 10,410 episodes for each noise level using each treatment modality with and without trend arrows, for a total of 437,220 simulations.

Outcome measures

The effect of POC use and of threshold and predictive hypoglycemia alarms was assessed using the frequency of moderate hypoglycemia below 50 mg/dL and of “severe” hypoglycemic episodes of ≤ 39 mg/dL. The effect of POC use and of threshold and predictive hyperglycemia alarms was assessed by the frequency of moderate hyperglycemia above 250 mg/dL and by the frequency of significant hyperglycemic episodes above 400 mg/dL. Average glucose was computed as well to assess overall BG control.

The magnitude of sensor error at which a significant deterioration of glucose control becomes observable was gauged by the inflection points observed in the graph—as commonly accepted, an abrupt change in the trend of the observed deterioration of any of the above metrics indicates a threshold beyond which sensor use for a particular treatment modality was no longer acceptable.

Results

Throughout this section we assess glycemic outcomes resulting from the use of three treatment modalities—POC, Threshold, and Predictive Alarm, each considered without and with the influence of trend arrows on the patient decision. Thus, all figures display six distinct traces. Each trace is plotted twice against sensor MARD on the x -axis or against the percentage of sensor errors exceeding 20%. The latter is

roughly equivalent to the percentage of sensor errors outside of Zone A of the Clarke or Parkes error-grid analysis—widely accepted tools for judging the clinical accuracy of SMBG and CGM devices.^{56,57}

Risk for hypoglycemia

Figure 1A and B presents the percentage time below 50 mg/dL for the six treatment modalities defined above (i.e., the time spent in moderate hypoglycemia). Figure 1C and D presents the number of hypoglycemic events ≤ 39 mg/dL for a 12-h period that required one or more treatments with 30 g of carbohydrate for recovery from hypoglycemia (i.e., the equivalent of biochemical severe hypoglycemic episodes). In Figure 1A and C the outcomes are plotted against MARD ranging from 3% to 22%; in Figure 1B and D the outcomes are plotted against the percentage of sensor errors exceeding 20%. Despite the difference in outcome due to treatment modality, a common pattern emerged: the trend lines describing the increase in the risk for hypoglycemia with increasing MARD changed their slope abruptly at approximately MARD=10% (Fig. 1A and C). This indicates that 10% is a natural cutoff value for sensor accuracy, below which the risk for hypoglycemia is relatively flat and above which the risk for hypoglycemia increases more rapidly, regardless of the treatment modality. No similar cutoff was observed along the percentage of sensor errors exceeding 20%—the increase in hypoglycemia risk there was relatively gradual (Fig. 1B and D).

Risk for hyperglycemia

Figure 2A and B presents the percentage time above 250 mg/dL for the six treatment modalities defined above (i.e., the time spent in moderate hyperglycemia). Figure 2C and D present the number of hyperglycemic events exceeding 400 mg/dL for a 12-h period (i.e., the equivalent of biochemical severe hyperglycemic episodes). The same common pattern as in Figure 1 emerges: the trend lines describing the increase in the risk for hyperglycemia with increasing MARD change their slope abruptly at approximately MARD=10% (Fig. 2A and C). This confirms that 10% is a cutoff value for sensor accuracy below which the risk for hyperglycemia is relatively flat and above which the risk increases more rapidly regardless of the treatment modality. The increase in risk for hyperglycemia plotted against sensor errors exceeding 20% is relatively gradual (Fig. 2B and D).

Average glucose control

Figure 3 presents average BG achieved in the six treatment scenarios defined above. Again, regardless of treatment modality, average BG stayed relatively flat under MARD=10%, after which average glucose control deteriorated rapidly. Here we also see that the influence of CGM trend arrow display is substantial—much greater than the influence of trend arrows on the extreme glycemic outcomes presented in Figures 1 and 2.

The mean improvement in average glycemia resulting from the use of additional CGM-specific information sources was 1.86 mg/dL due to the use of trend arrows, 8.17 mg/dL due to the use of threshold alarms, and 8.88 mg/dL due to the use of predictive alarms.

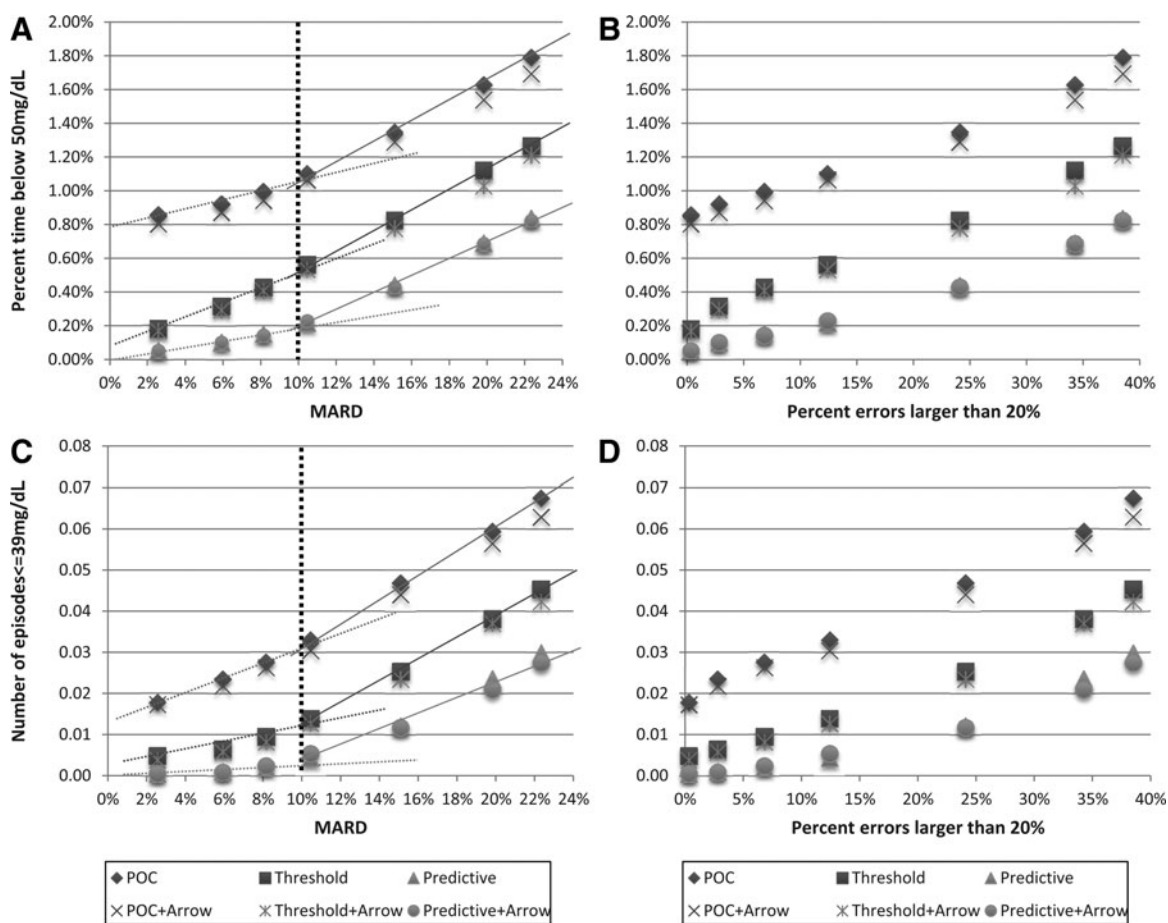


FIG. 1. Hypoglycemia outcomes from six treatment modalities plotted (A and C) against sensor mean absolute relative difference (MARD) and (B and D) against the frequency of sensor errors exceeding 20%: (A and B) percentage of blood glucose values below 50 mg/dL and (C and D) frequency of blood glucose readings ≤ 39 mg/dL within a 12-h time period. POC, point-of-care treatment equivalent to direct dosing using a sensor value; Threshold, addition of threshold alarms to prevent hypoglycemia; Predictive, addition of predictive alarms to prevent hypoglycemia. POC + Arrow, Threshold + Arrow, and Predictive + Arrow are settings that use additional insulin reduction based on downward trend arrows displayed by the sensor.

Sensor error characteristics

Last, but not least, the distribution of CGM errors used in this study might be of additional interest. Figure 4 presents the dependence between MARD and the frequency of large sensor deviations (e.g., errors greater than 20% and errors greater than 50%) (Fig. 4A) as well as the bias and the quartiles of the CGM error distribution (Fig. 4B).

Discussion

Because CGM errors and deviations are inherent to the physiology of CGM and unlikely to be completely eliminated, an overarching question is to what extent CGM errors influence the quality of diabetes control? In other words, what degree of accuracy would allow CGM to be the primary source of information for diabetes treatment (i.e., a replacement for SMBG). The goal of this study was to answer this question via combination of real-time observation of patients with type 1 diabetes over a month and a series of computer simulation experiments aiming to identify CGM accuracy levels in the context of various diabetes treatment modalities. Although we have done simulation experiments linking SMBG accuracy to

glycemic outcomes,⁵⁸ to the best of our knowledge, a combination of in vivo and in silico experiments has not been attempted before. Such a combination allowed real treatment outcomes observed in the field to be “replayed” in computer simulation with precisely manipulated CGM error—an experiment that is evidently impossible in clinical studies.

To do so, we used our archival database containing approximately 3 patient-years of CGM and insulin pump data, accompanied by SMBG readings and records of meals and other treatment-related events. Using these data and a mathematical model of the human metabolic system in diabetes, we reconstructed the nominal “true” glucose profile for each person. Then we created a realistic in silico description of sensor errors, which recreated the temporal structure of sensor deviations observed in vivo. This allowed the nominal traces observed in the field to be “replayed” with increasing error magnitude. We repeated this procedure with a basic POC dosing of insulin using CGM values (e.g., a direct replacement of SMBG by CGM) and with several other treatment modalities that use the rich information available in the CGM data stream (but inaccessible to SMBG): trend arrows, threshold, and predictive alarms.

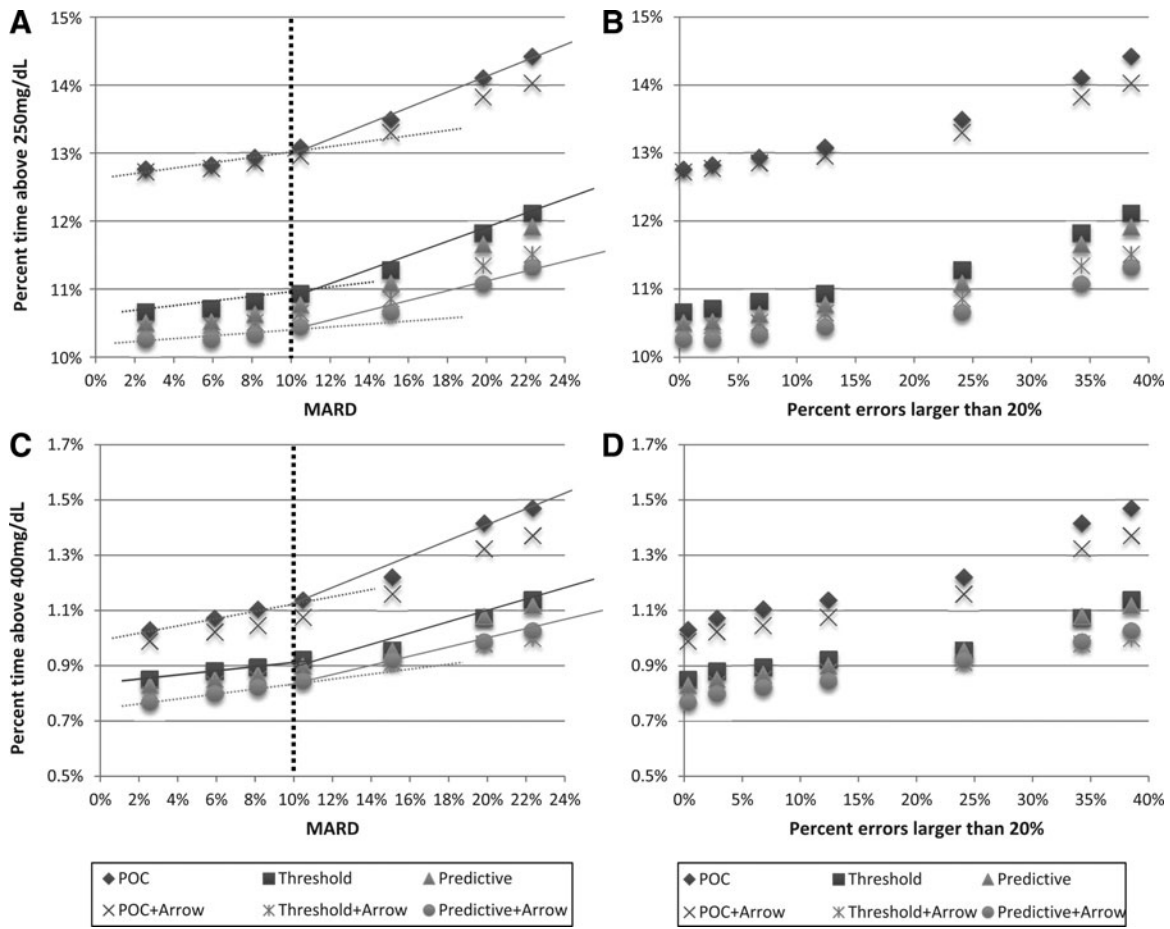


FIG. 2. Hyperglycemic outcomes from six treatment modalities plotted (A and C) against sensor mean absolute relative difference (MARD) and (B and D) against the frequency of sensor errors exceeding 20%: (A and B) percentage of blood glucose values above 250 mg/dL and (C and D) percentage of blood glucose readings above 400 mg/dL. POC, point-of-care treatment equivalent to direct dosing using a sensor value; Threshold, addition of threshold alarms to prevent hyperglycemia; Predictive, addition of predictive alarms to prevent hyperglycemia. POC+ Arrow, Threshold+ Arrow, and Predictive+ Arrow are settings that use additional insulin increase based on upward trend arrows displayed by the sensor.

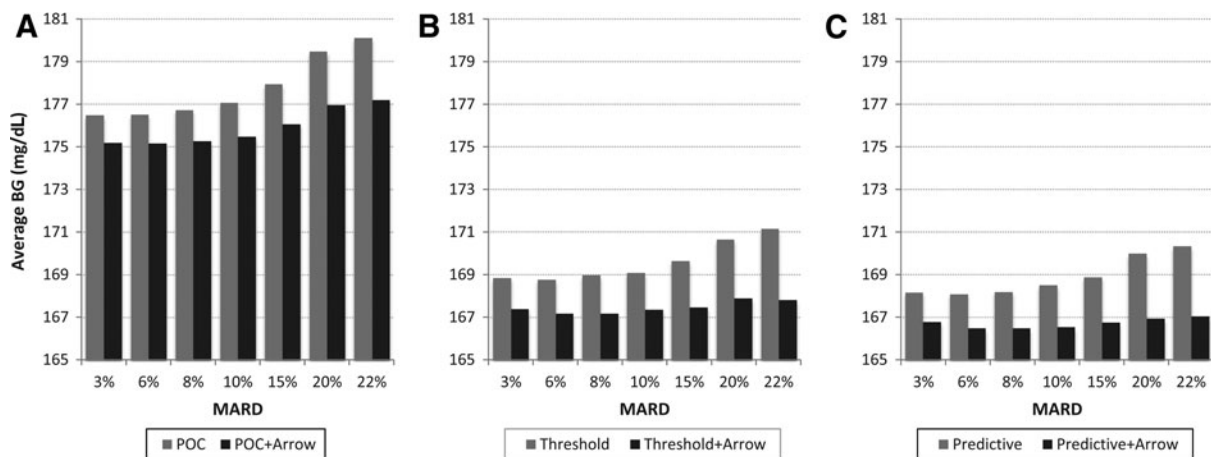


FIG. 3. Average blood glucose (BG) levels resulting from six treatment modalities plotted against sensor mean absolute relative difference (MARD): (A) point-of-care (POC) treatment equivalent to direct dosing using a sensor value; (B) the effect of threshold alarms; and (C) the effect of predictive alarms. All treatment modalities were tested twice, without and with additional insulin increase based on upward trend arrows displayed by the sensor.

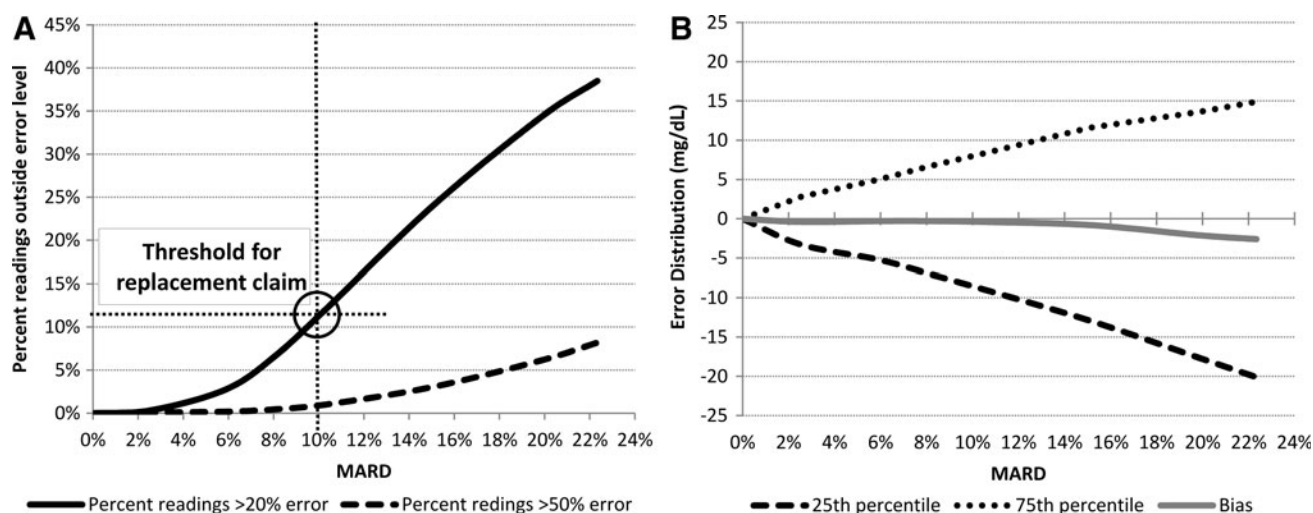


FIG. 4. Characteristics of the distribution of continuous glucose monitoring errors implemented in this study plotted by mean absolute relative difference (MARD): (A) percentage errors greater than 20% and 50% and (B) quartiles and bias of the error distribution in mg/dL.

In terms of outcomes, we limited our output to established metrics of glycemic control, such as percentage of time above/below glycemic thresholds and average glucose. We intentionally omitted intermediate metrics such as hit/false alarm rate of hypoglycemia detection because such metrics have been controversial in the past, and their inclusion could have diverted the discussion from the primary goal of this study—to establish a threshold for sensor accuracy that would permit non-adjunct CGM use as a replacement for SMBG appropriate for insulin dosing.

The major limitation of this study is the use of *in silico* experiments to predict treatment outcomes. We should re-emphasize, however, that finding the error thresholds below which a CGM replacement claim is feasible and safe is not possible in a clinical trial because the exact same glycemic/treatment conditions cannot be reproduced *in vivo* at varying degrees of sensor error. Thus, conceptually, the only option we have to gauge the degree of influence of sensor error on glycemic control is modeling and simulation, such as those presented in this article. This said, we also grounded our analyses on over 2,000 segments of real-life treatment patterns, and the only variable we manipulated was the magnitude of sensor errors. Therefore, the presented results should be sufficiently credible to be used as a guideline for future clinical trials and technology development targets.

We should also emphasize that the only criterion we used to determine a threshold value for sensor accuracy below which further improvement in MARD contributes little to better outcomes was the slope change of risk increase in various treatment modalities. We do not use any artificial thresholds of risk for hypo- or hyperglycemia below which we would claim that the risks were insignificant (e.g., we do not assert that 1% time in hypoglycemia below 50 mg/dL is acceptable or 0.025 extreme hypoglycemic episodes per 12 h is low enough). Such assertions can be clinically controversial and are not needed for our conclusions. With this in mind, we would like to avoid a possible misconception of the presented results: in all figures, the use of threshold or predictive alarms results in substantially better clinical outcomes than

the use of direct sensor readings alone. *This does not mean that alarms can be used as a substitute for accuracy, even if better outcomes are achieved with the aid of an alarm at higher MARD.* This is because alarms are not effective 100% of the time and are not always heard or reacted to properly and thus can be a barrier to CGM acceptance.⁵⁹ For example, if half of the threshold alarms were missed in Figure 1A, the threshold alarm line will move half-way up towards the POC line. Therefore, a sensor should be judged as a possible replacement for SMBG on its basic accuracy in a direct POC treatment modality.

Along these lines, an interesting observation can be made when comparing the effects of predictive versus threshold alarms in Figures 1–3: In Figure 1 this effect is relatively large, which is to be expected—a predictive alarm prompting a fast-acting treatment should work better for avoidance of hypoglycemia than an alarm issued when a hypoglycemic threshold is already reached. In contrast, in Figures 2 and 3, there is little difference between the effects of predictive and threshold alarms; this is expected as well—owing to insulin action time lag, the mitigation of hyperglycemia (and therefore the effect on average glucose) would be relatively minor with 15–20 min of predictive action.

In addition to average accuracy as represented by MARD, sensors should be evaluated for their ability to replace SMBG on other parameters as well. Most important is that CGM *reliability* must accompany CGM accuracy—having a 10% MARD would be sufficient for non-adjunctive use only if the sensor produces reliable data without signal interruption or loss of sensitivity. Reliability must be sustained over the claimed sensor life and must be consistent from one sensor to the next. The frequency of outliers (e.g., readings that have errors of 20% or more) can impact treatment decisions as well, particularly if the deviations persist over time. As seen in Figures 1B–D and 2B–D, outliers influence outcomes almost gradually (with a possible inflection point between 10% and 15%); thus outlier frequency should be mitigated as much as technologically possible. Future studies should therefore look at factors such as out-the-box failures, sensor

longevity with respect to intended use, sensor-to-sensor consistency, and frequency of outliers. In particular, the clinical interpretation of sensor errors and outliers would benefit from the recently introduced Surveillance Error-Grid, which was exclusively designed to gauge the clinical impact of SMBG errors, but could be adapted to evaluation of CGM error patterns.^{60,61}

In conclusion, CGM errors, time lag, and transient loss of sensitivity are primary obstacles to the adoption of this technology as a replacement to “fingerstick” self-monitoring. The thesis of this article is that CGM data contain rich information about a person’s glycemic state and can help improve glycemic control if the information is processed appropriately. Our primary conclusion is that direct use of CGM for insulin dosing decisions is feasible below a certain level of sensor error, estimated by our experiments at MARD=10%. This level of overall accuracy should be accompanied by minimizing the frequency of large sensor deviations (e.g., those >20%), which can additionally contribute to poorer glycemic outcomes. In our experiments, further increase of sensor accuracy below the 10% MARD threshold did not contribute substantively to better glycemic outcomes.

Acknowledgments

The *in silico* trials and the writing of this manuscript were supported by a research grant from Becton, Dickinson, and Company to the University of Virginia. The authors thank Dr. James Petisce, BD Diabetes Care, for his thoughtful comments and for his help with the validation of our results. The database used in this study was built during Phase 1 of NIH/NIDDK project RO1 DK 085623. The JDRF Artificial Pancreas Project at the University of Virginia supported the building of the simulation environment used in this study. The authors thank Insulet Corp. for providing insulin pumps and Dexcom Inc. for providing CGM sensors used for the building of this database.

Author Disclosure Statement

B.P.K. has served on the advisory panels of Animas Corporation and Sanofi-Aventis and has received research support from Amylin Pharmaceuticals, Animas, Dexcom, Insulet, Roche Diagnostics, Sanofi-Aventis, and Tandem Diabetes Care. M.D.B. has received honoraria from Bayer, Roche Diagnostics, and Sanofi-Aventis and research support from Animas, Dexcom, Insulet, Sanofi-Aventis, and Tandem Diabetes Care. B.P.K., S.D.P., and M.D.B. hold patents and patent applications related to SMBG and CGM technologies. E.A.O. declares no competing financial interests exist.

References

- Mastrototaro JJ: The MiniMed continuous glucose monitoring system. *Diabetes Technol Ther* 2000;2(Suppl 1):S-13–S-18.
- Bode BW: Clinical utility of the continuous glucose monitoring system. *Diabetes Technol Ther* 2000;2(Suppl 1):S-35–S-42.
- Feldman B, Brazg R, Schwartz S, Weinstein R: A continuous glucose sensor based on wired enzyme technology—results from a 3-day trial in patients with type 1 diabetes. *Diabetes Technol Ther* 2003;5:769–778.
- Deiss D, Bolinder J, Riveline J, Battelino T, Bosi E, Tubiana-Rufi N, Kerr D, Phillip M: Improved glycemic control in poorly controlled patients with type 1 diabetes using real-time continuous glucose monitoring. *Diabetes Care* 2006;29:2730–2732.
- Garg K, Zisser H, Schwartz S, Bailey T, Kaplan R, Ellis S, Jovanovic L: Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor. *Diabetes Care* 2006;29:44–50.
- Kovatchev BP, Clarke WL: Continuous glucose monitoring reduces risks for hypo- and hyperglycemia and glucose variability in diabetes [abstract]. *Diabetes* 2007;56(Suppl 1):0086OR.
- The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group: Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med* 2008;359:1464–1476.
- Klonoff DC: Continuous glucose monitoring: roadmap for 21st century diabetes therapy. *Diabetes Care* 2005;28:1231–1239.
- Hirsch IB, Armstrong D, Bergenstal RM, Buckingham B, Childs BP, Clarke WL, Peters A, Wolpert H: Clinical application of emerging sensor technologies in diabetes management: consensus guidelines for continuous glucose monitoring. *Diabetes Technol Ther* 2008;10:232–246.
- Hovorka R: Continuous glucose monitoring and closed-loop systems. *Diabet Med* 2006;23:1–12.
- Klonoff DC: The artificial pancreas: how sweet engineering will solve bitter problems. *J Diabetes Sci Technol* 2007;1:72–81.
- Rebrin K, Steil GM, van Antwerp WP, Mastrototaro JJ: Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring. *Am J Physiol Endocrinol Metab* 1999;277:E561–E571.
- Aussedat B, Dupire-Angel M, Gifford R, Klein JC, Wilson GS, Reach G: Interstitial glucose concentration and glycemia: implications for continuous subcutaneous glucose monitoring. *Am J Physiol Endocrinol Metab* 2000;278:E716–E728.
- Steil GM, Rebrin K, Hariri F, Jinagonda S, Tadros S, Darwin C, Saad MF: Interstitial fluid glucose dynamics during insulin-induced hypoglycaemia. *Diabetologia* 2005;48:1833–1840.
- Boyne M, Silver D, Kaplan J, Saudek C: Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. *Diabetes* 2003;52:2790–2794.
- Kulcu E, Tamada JA, Reach G, Potts RO, Lesho MJ: Physiological differences between interstitial glucose and blood glucose measured in human subjects. *Diabetes Care* 2003;26:2405–2409.
- Wientjes KJ, Schoonen AJ: Determination of time delay between blood and interstitial adipose tissue glucose concentration change by microdialysis in healthy volunteers. *Int J Artif Organs* 2001;24:884–889.
- Wentholt IME, Hart AAM, Hoekstra JBL, DeVries JH: Relationship between interstitial and blood glucose in type 1 diabetes patients: delay and the push-pull phenomenon revisited. *Diabetes Technol Ther* 2004;9:169–175.
- Basu A, Dube S, Slama M, Errazuriz I, Amezcua JC, Kudva YC, Peyser T, Carter RE, Cobelli C, Basu R: Time lag of glucose from intravascular to interstitial compartment in humans. *Diabetes* 2013;62:4083–4087.

20. Kovatchev BP, Clarke WL: Peculiarities of the continuous glucose monitoring data stream and their impact on developing closed-loop control technology. *J Diabetes Sci Technol* 2008;2:158–163.
21. Calhoun P, Lum J, Beck RW, Kollman C: Performance comparison of the Medtronic Sof-Sensor and Enlite glucose sensors in inpatient studies of individuals with type 1 diabetes. *Diabetes Technol Ther* 2013;15:758–761.
22. Christiansen M, Bailey T, Watkins E, Liljenquist D, Price D, Nakamura K, Boock R, Peyser T: A new-generation continuous glucose monitoring system: improved accuracy and reliability compared with a previous-generation system. *Diabetes Technol Ther* 2013;15:881–888.
23. Freckmann G, Pleus S, Link M, Zschornack E, Klötzer H-M, Haug C: Performance evaluation of three continuous glucose monitoring systems: comparison of six sensors per subject in parallel. *J Diabetes Sci Technol* 2013;7:842–853.
24. Pleus S, Schmid C, Link M, Zschornack E, Klötzer H-M, Haug C, Freckmann G: Performance evaluation of a continuous glucose monitoring system under conditions similar to daily life. *J Diabetes Sci Technol* 2013;7:833–841.
25. Bailey TS, Ahmann A, Brazg R, Christiansen M, Garg S, Watkins E, Welsh JB, Lee SW: Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. *Diabetes Technol Ther* 2014;16:277–283.
26. Kropff J, Bruttomesso D, Doll W, Farret A, Galasso S, Luijff YM, Mader JK, Place J, Boscari F, Pieber TR, Renard E, DeVries JH: Accuracy of two continuous glucose monitoring systems: a head-to-head comparison under clinical research centre and daily life conditions. *Diabetes Obes Metab* 2014 August 11 [Epub ahead of print]. doi: 10.1111/dom.12378.
27. Damiano ER, McKeon K, El-Khatib FH, Zheng H, Nathan DM, Russell SJ: A comparative effectiveness analysis of three continuous glucose monitors: the Navigator, G4 Platinum, and Enlite. *J Diabetes Sci Technol* 2014 April 21 [Epub ahead of print]. DOI: 10.1177/1932296814532203
28. Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL: Evaluating the accuracy of continuous glucose monitoring sensors: continuous glucose error-grid analysis (CG-EGA) illustrated by Therasense Freestyle Navigator™ data. *Diabetes Care* 2004;27:1922–1928.
29. Clinical and Laboratory Standards Institute (CLSI): Performance Metrics for Continuous Interstitial Glucose Monitoring; Approved Guideline. CLSI document POCT05-A. Wayne, PA: CLSI, 2008.
30. McGarraugh GV, Clarke WL, Kovatchev BP: Comparison of the clinical information provided by the FreeStyle Navigator continuous interstitial glucose monitor versus traditional blood glucose readings. *Diabetes Technol Ther* 2010;12:365–371.
31. Clarke WL, Kovatchev BP: Statistical tools to analyze CGM data. *Diabetes Technol Ther* 2009;11(Suppl 1):S-45–S-54.
32. Cobelli C, Dalla Man C, Sparacino G, Magni L, Nicolao G, Kovatchev BP: Diabetes: models, signals, and control. *IEEE Rev Biomed Eng* 2009;2:54–96.
33. Davis SN, Horton ES, Battelino T, Rubin RR, Schulman KA, Tamborlane WV: STAR 3 randomized controlled trial to compare sensor-augmented insulin pump therapy with multiple daily injections in the treatment of type 1 diabetes: research design, methods, and baseline characteristics of enrolled subjects. *Diabetes Technol Ther* 2010;12:249–255.
34. Battelino T, Conget I, Olsen B, Schütz-Fuhrmann I, Hommel E, Hoogma R, Schierloh U, Sulli N, Bolinder J; SWITCH Study Group: The use and efficacy of continuous glucose monitoring in type 1 diabetes treated with insulin pump therapy: a randomised controlled trial. *Diabetologia* 2012;55:1355–1362.
35. Heise T, Koschinsky T, Heinemann L, Lodwig V; Glucose Monitoring Study Group: Hypoglycemia warning signal and glucose sensors: requirements and concepts. *Diabetes Technol Ther* 2003;5:563–571.
36. Choudhary P, Shin J, Wang Y, Evans ML, Hammond PJ, Kerr D, Shaw JA, Pickup JC, Amiel SA: Insulin pump therapy with automated insulin suspension in response to hypoglycemia: reduction in nocturnal hypoglycemia in those at greatest risk. *Diabetes Care* 2011;34:2023–2025.
37. Danne T, Kordonouri O, Holder M, Haberland H, Golembowski S, Remus K, Bläsing S, Wadien T, Zierow S, Hartmann R, Thomas A: Prevention of hypoglycemia by using low glucose suspend function in sensor-augmented pump therapy. *Diabetes Technol Ther* 2011;13:1129–1134.
38. Bergenstal RM, Klonoff DC, Garg SK, Bode BW, Meredith M, Slover RH, Ahmann AJ, Welsh JB, Lee SW, Kaufman FR; ASPIRE In-Home Study Group: Threshold-based insulin-pump interruption for reduction of hypoglycemia. *N Engl J Med* 2013;369:224–232.
39. Buckingham B, Cobry E, Clinton P, Gage V, Caswell K, Kunselman E, Cameron F, Chase HP: Preventing hypoglycemia using predictive alarm algorithms and insulin pump suspension. *Diabetes Technol Ther* 2009;11:93–97.
40. Weinzimer SA, Steil GM, Swan KL, Dziura J, Kurtz N, Tamborlane WV: Fully automated closed-loop insulin delivery versus semi-automated hybrid control in pediatric patients with type 1 diabetes using an artificial pancreas. *Diabetes Care* 2008;31:934–939.
41. Clarke WL, Anderson SM, Breton MD, Patek SD, Kashmer L, Kovatchev BP: Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: the Virginia experience. *J Diabetes Sci Technol* 2009;3:1031–1038.
42. Hovorka R, Allen JM, Elleri D, Chassin LJ, Harris J, Xing D, Kollman C, Hovorka T, Larsen AM, Nodale M, De Palma A, Wilinska ME, Acerini CL, Dunger DB: Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. *Lancet* 2010;375:743–751.
43. El-Khatib FH, Russell SJ, Nathan DM, Sutherland RG, Damiano ER: A bihormonal closed-loop artificial pancreas for type 1 diabetes. *Science Transl Med* 2010;2:27ra27.
44. Atlas E, Nimri R, Miller S, Grunberg EA, Phillip M: MD-Logic artificial pancreas system: a pilot study in adults with type 1 diabetes. *Diabetes Care* 2010;33:1072–1076.
45. Hovorka R, Kumareswaran K, Harris J, Allen JM, Elleri D, Xing D, Kollman C, Nodale M, Murphy HR, Dunger DB, Amiel SA, Heller SR, Wilinska ME, Evans ML: Overnight closed loop insulin delivery in adults with type 1 diabetes: crossover randomised controlled studies. *BMJ* 2011;342:d1855.
46. Cobelli C, Renard E, Kovatchev BP: Artificial pancreas: past, present, future. *Diabetes* 2011;60:2672–2682.
47. Kovatchev BP, Renard E, Cobelli C, Zisser H, Keith-Hynes P, Anderson SM, Brown SA, Chernavvsky DR, Breton MD, Farret A, Pelletier MJ, Place J, Bruttomesso D, Del Favero S, Visentin R, Filippi A, Scotton R, Avogaro A, Doyle FJ III: Feasibility of outpatient fully integrated

- closed-loop control: first studies of wearable artificial pancreas. *Diabetes Care* 2013;36:1851–1858.
48. Phillip M, Battelino T, Atlas E, Kordonouri O, Bratina N, Miller S, Biester T, Stefanija MA, Muller I, Nimri R, Danne T: Nocturnal glucose control with an artificial pancreas at a diabetes camp. *N Engl J Med* 2013;368:824–833.
 49. Nimri R, Muller I, Atlas E, Miller S, Kordonouri O, Bratina N, Tsioli C, Stefanija MA, Danne T, Battelino T, Phillip M: Night glucose control with MD-Logic artificial pancreas in home setting: a single blind, randomized crossover trial-interim analysis. *Pediatr Diabetes* 2014;15:91–99.
 50. Renard E, Cobelli C, Kovatchev BP: Closed loop developments to improve glucose control at home. *Diabetes Res Clin Pract* 2013;102:79–85.
 51. Kovatchev BP, Renard E, Cobelli C, Zisser H, Keith-Hynes P, Anderson SM, Brown SA, Chernavvsky DR, Breton MD, Mize LB, Farret A, Place J, Bruttomesso D, Del Favero S, Boscari F, Galasso S, Avogaro A, Magni L, Di Palma F, Toffanin C, Messori M, Dassay E, Doyle F III: Safety of outpatient closed-loop control: first randomized crossover trials of a wearable artificial pancreas. *Diabetes Care* 2014;37:1789–1796.
 52. Shepard JA, Gonder-Frederick LA, Vajda K, Kovatchev BP: Patient perspectives on personalized glucose advisory systems for type 1 diabetes management. *Diabetes Technol Ther* 2012;14:858–861.
 53. Beck RW, Calhoun P, Kollman C: Challenges for outpatient closed loop studies: how to assess efficacy. *Diabetes Technol Ther* 2013;15:1–3.
 54. Hughes CS, Patek SD, Breton MD, Kovatchev BP: Hypoglycemia prevention via pump attenuation and red-yellow-green “traffic” lights using continuous glucose monitoring and insulin pump data. *Diabetes Sci Technol* 2010;4:1146–1155.
 55. Breton MD, Kovatchev BP: Analysis, modeling, and simulation of the accuracy of continuous glucose sensors. *J Diabetes Sci Technol* 2008;2:853–862.
 56. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL: Evaluating the clinical accuracy of self-blood glucose monitoring systems. *Diabetes Care* 1987;10:622–628.
 57. Parkes JL, Slatin SL, Pardo S, Ginsberg BH: A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care* 2000;23:1143–1148.
 58. Breton MD, Kovatchev BP: Impact of blood glucose self-monitoring errors on glucose variability, risk for hypoglycemia, and average glucose control in type 1 diabetes: an in silico study. *J Diabetes Sci Technol* 2010;4:562–570.
 59. Shivers JP, Mackowiak L, Anhalt H, Zisser H: “Turn it off!”: diabetes device alarm fatigue considerations for the present and the future. *J Diabetes Sci Technol* 2013;7:789–794.
 60. Klonoff DC, Lias C, Vigersky R, Clarke WL, Parkes JL, Sacks DB, Kirkman MS, Kovatchev BP: Error Grid Panel: The surveillance error grid. *J Diabetes Sci Technol* 2014;8:658–672.
 61. Kovatchev BP, Wakeman CA, Breton MD, Kost GJ, Louie RF, Tran NK, Klonoff DC: Computing the surveillance error grid analysis: procedure and examples. *J Diabetes Sci Technol* 2014;8:673–684.

Address correspondence to:

Boris Kovatchev, PhD

University of Virginia Health System

P.O. Box 400888

Charlottesville, VA 22908

E-mail: boris@virginia.edu