

ORIGINAL ARTICLE

Clinically Relevant Hypoglycemia Prediction Metrics for Event Mitigation

Rebecca A. Harvey, B.S.,^{1,2} Eyal Dassau, Ph.D.,^{1,2,3} Howard C. Zisser, M.D.,^{1,2} Wendy Bevier, Ph.D.,¹ Dale E. Seborg, Ph.D.,^{1,2} Lois Jovanovič, M.D.,^{1,2,3} and Francis J. Doyle III, Ph.D.^{1,2,3}

Abstract

Background: The purpose of this study was to develop a method to compare hypoglycemia prediction algorithms and choose parameter settings for different applications, such as triggering insulin pump suspension or alerting for rescue carbohydrate treatment.

Materials and Methods: Hypoglycemia prediction algorithms with different parameter settings were implemented on an ambulatory dataset containing 490 days from 30 subjects with type 1 diabetes mellitus using the Dexcom™ (San Diego, CA) SEVEN™ continuous glucose monitoring system. The performance was evaluated using a proposed set of metrics representing the true-positive ratio, false-positive rate, and distribution of warning times. A prospective, in silico study was performed to show the effect of using different parameter settings to prevent or rescue from hypoglycemia.

Results: The retrospective study results suggest the parameter settings for different methods of hypoglycemia mitigation. When rescue carbohydrates are used, a high true-positive ratio, a minimal false-positive rate, and alarms with short warning time are desired. These objectives were met with a 30-min prediction horizon and two successive flags required to alarm: 78% of events were detected with 3.0 false alarms/day and 66% probability of alarms occurring within 30 min of the event. This parameter setting selection was confirmed in silico: treating with rescue carbohydrates reduced the duration of hypoglycemia from 14.9% to 0.5%. However, for a different method, such as pump suspension, this parameter setting only reduced hypoglycemia to 8.7%, as can be expected by the low probability of alarming more than 30 min ahead.

Conclusions: The proposed metrics allow direct comparison of hypoglycemia prediction algorithms and selection of parameter settings for different types of hypoglycemia mitigation, as shown in the prospective in silico study in which hypoglycemia was alerted or treated with rescue carbohydrates.

Background

INTENSIVE INSULIN THERAPY is recommended to reduce the likelihood of long-term complications for those with type 1 diabetes mellitus. Unfortunately, this therapy results in a “two-to-threefold increase in severe hypoglycemia,” as found in the landmark study performed by the Diabetes Control and Complications Trial.¹ Frequent hypoglycemia episodes can lead to hypoglycemia unawareness, which causes symptoms such as confusion, sweating, and dizziness to occur at lower blood glucose levels.² Although the sympathetic response to hypoglycemia varies, the American Diabetes Association recommends an alert level of 70 mg/dL (3.9 mmol/L)³; this value will be used throughout this article, but is a flexible parameter. The availability and increasing accuracy of continuous glucose monitors (CGMs) have allowed for better glycemic control and the possibility of prediction of adverse

glycemic excursions.^{4–6} The potential for overnight severe hypoglycemia and the absence of symptoms for many with type 1 diabetes must be addressed in current CGMs and eventually in the development of an artificial pancreas. A prediction algorithm used for this purpose must be able to accurately predict hypoglycemia episodes and attempt to mitigate them before they occur.^{2,7}

Several promising hypoglycemia prediction algorithms have been described in the literature.^{4,5,8–10} However, there is a lack of consistency, both in the type of data used to assess the algorithms and in the reporting of their performance, making it difficult to compare them. Several of the algorithms were assessed using data in which hypoglycemia was induced,^{4,5,8,9} although, in practice, severe hypoglycemia is relatively infrequent, and the drop in blood glucose in an induced study may exhibit an unrealistic pattern. Other algorithms were tested on data that had been post-processed by

¹Sansum Diabetes Research Institute, Santa Barbara, California.

²Department of Chemical Engineering and ³Biomolecular Science & Engineering Program, University of California, Santa Barbara, Santa Barbara, California.

calibration with fingerstick data, whereas in real time this would not be feasible.¹⁰ Assessing a prediction algorithm with long-term ambulatory conditions gives insight into its performance in real time, particularly the rate of false-positive alarms during long data segments without hypoglycemia episodes.

The performance necessary to determine the best algorithm tuning depends on the method of action because the timing of the alarms and the specificity and sensitivity required vary depending on the treatment to be used. For example, preventing hypoglycemia by suspending the insulin pump is a type of passive protection that requires a warning time of approximately 45 min to be successful.⁴ Current CGMs are equipped with variable threshold alarm settings, which are useful, but limited, as a high threshold may lead to many false-positive alarms, and a low threshold may not allow for adequate time for corrective action. A predictive alarm can exploit the recent data trends to predict hypoglycemia more accurately and earlier. The initial step for bringing the artificial pancreas to those with type 1 diabetes will be the use of pump suspension with a threshold or predictive alarm,^{11,12} but this may lead to rebound hyperglycemia with false-positive alarms or may be unsuccessful if triggered too late.¹³ Therefore, an algorithm designed for this use must have a long warning time and a high true-positive ratio with a moderate to low false-positive rate. Alternatively, alerting to the user to treat with rescue carbohydrates (CHOs) is a type of active protection that should occur very close to the event. Because the current CGMs are somewhat noisy (mean absolute deviation of 2.6–22.6 mg/dL [0.14–1.26 mmol/L] below 70 mg/dL [3.9 mmol/L]),¹⁴ false-positive alarms may be frequent and can lead to alarm fatigue with audible alarms.¹⁵ This method of action requires an algorithm tuned for a short warning time with a low false-positive rate.

In light of these requirements, an algorithm must be designed to produce alarms with appropriate true- and false-positive ratios and a warning time that allows corrective action to be taken. We have developed metrics to address these goals that include a practical calculation of true- and false-positive rates and the representation of warning times as a distribution, in effect visualizing both clinical and engineering requirements. These metrics were applied to a large ambulatory dataset to determine the performance of different algorithms and parameter settings, and a prospective *in silico* study was performed using a variety of parameter settings to demonstrate the impact of algorithm tuning on hypoglycemia prevention or mitigation.

Materials and Methods

Prediction algorithms

The proposed metrics were demonstrated using a simple numerical logic algorithm (NLA), similar to one that has been used in previous studies.^{4,8} In the NLA, the rate of change was calculated at each point using the first derivative of the Lagrange interpolation polynomial (which is robust to differing sampling rates) to make a prediction and to issue an alarm if a hypoglycemic event was imminent.^{8,16} The trajectory of the rate of change estimate was projected through the hypoglycemia threshold to decide if a hypoglycemic event would occur within the prediction horizon (*PH*), the main tuning parameter.

The influence of sensor noise was reduced by using an additional parameter, the number of successive alarms required (*SAR*) to register an alarm; an *SAR* value of either 1 or 2 was used in this study. If a hypoglycemic event was predicted at *SAR* successive measurements, an alarm was recorded.

An additional algorithm using the Kalman filter was implemented to illustrate that these metrics can be used with any algorithm. The algorithm used was a modified version of the Optimal Estimation (OE) algorithm, introduced by Palerm et al.,¹⁷ in which a Kalman filter was used to estimate glucose concentration and its rate of change. These estimates were used to make glucose predictions. The algorithm was modified to allow for missing data by using the most recent prediction to fill in gaps up to 15 min (the sampling period was 5 min). When longer gaps were encountered, the algorithm was restarted. The tuning parameters used for the OE algorithm were the values of *Q* and *R* in the quadratic performance index. In this study, the *Q/R* ratio was set at 0.008. For more information about the Kalman filter algorithm, see Palerm et al.¹⁷

Retrospective clinical study

The retrospective study was based on 38 sets of ambulatory data from 30 subjects with type 1 diabetes mellitus (negative C-peptide concentration), collected by the Sansum Diabetes Research Institute, Santa Barbara, CA. The record consisted of 490 days of CGM data (SEVEN™ CGM; Dexcom™, San Diego, CA) with a 5-min sampling time. These data represent a real-life record in which hypoglycemia occurs with a typical frequency under ambulatory conditions. For further information, see Table 1. The data were processed with three *PH* values (15, 30, and 60 min) for NLA and OE and two *SAR* values (1 and 2 successive alarms required) for NLA.

In silico study

An *in silico* study to mimic hypoglycemia caused by an over-bolus was conducted using the Food and Drug Administration-accepted UVA/Padova metabolic simulator set of 10 published subjects.¹⁸ The simulated CGM, with preprogrammed noise, was used for all calculations. The 18-h simulation protocol was as follows: 2 h of basal insulin, followed by closed-loop therapy of basal insulin with a bolus for an anticipated meal of 65 g of CHO at 2.5 h, but without actual delivery of the meal. Eight scenarios were executed: one control with no bolus, one protocol as described above, three suspension scenarios, and three alert scenarios, using NLA.

For the suspension and alert scenarios, the NLA was active only during closed-loop. When the algorithm provided a prediction of future hypoglycemia, treatment was given, either to suspend the insulin pump to 0 U/h for 90 min (suspension scenario) or to administer 16 g of CHO as rescue CHOs (alert scenario). Further treatment was prevented for 30 min. The parameter settings were *SAR* = 2 and *PH* values of 15, 30, and 60 min.

Event determination

When a subject experiences a threshold low blood glucose alarm in daily life, he or she assesses the validity of an alarm based on CGM and fingerstick measurements. Because of this

TABLE 1. CHARACTERISTICS OF OVERALL CLINICAL DATA, THE DEMONSTRATIVE CASE, AND THE SIMULATION STUDY

	<i>All subjects</i>	<i>Demonstrative case</i>	<i>Simulation study</i>
Population data			
Number of datasets (distinct subjects)	38 (30)	1 (1)	10 (10)
Male sex [<i>n</i> (%)]	14 (47)	1 (100)	—
Age at collection (years) ^a	45 ± 14	62	—
BMI at collection (kg/m ²) ^a	26 ± 4	26	—
Sample period (min)	5	5	5
Overall duration [<i>n</i> samples (days)]	141,178 (490)	28,369 (98.5)	2,170 (7.5)
Median duration per subject [<i>n</i> samples (days)]	1,566 (5)	—	216 (0.75)
Below 70 mg/dL (3.9 mmol/L) [<i>n</i> samples (%)]	13,859 (9.8)	1,618 (5.7)	324 (14.9)
Above 180 mg/dL (10 mmol/L) [<i>n</i> samples (%)]	28,235 (20.0)	5,068 (17.9)	8 (0.4)
Hypoglycemia episodes [total (mean)]	789 (21)	83	10 (1)
False-positive range [<i>n</i> samples (% of total)]	111,848 (79.2)	24,867 (87.7)	1,635 (75.4)
Algorithm results			
<i>TPR</i> (number of episodes detected) (%)			
NLA, SAR = 1, PH = 15 min	86	87	100
NLA, SAR = 1, PH = 30 min	90	89	100
NLA, SAR = 1, PH = 60 min	92	95	100
NLA, SAR = 2, PH = 15 min	55	59	90
NLA, SAR = 2, PH = 30 min	75	78	100
NLA, SAR = 2, PH = 60 min	80	82	100
OE, PH = 15 min	57	70	80
OE, PH = 30 min	79	81	90
OE, PH = 60 min	89	90	90
<i>FPR</i> (false-positive alarms) (<i>n</i> alarms/day)			
NLA, SAR = 1, PH = 15 min	5.4	4.8	2.6
NLA, SAR = 1, PH = 30 min	12	11	11
NLA, SAR = 1, PH = 60 min	21	19	21
NLA, SAR = 2, PH = 15 min	1.1	0.9	1.1
NLA, SAR = 2, PH = 30 min	3.7	3.0	6.9
NLA, SAR = 2, PH = 60 min	8.3	6.8	16
OE, PH = 15 min	3.5	2.6	2.3
OE, PH = 30 min	8.8	7.1	5.6
OE, PH = 60 min	23	19	17
<i>t_w</i> for 60-min window (min) ^b			
NLA, SAR = 1, PH = 15 min	27 ± 16 (20)	26 ± 15 (20)	17 ± 8 (15)
NLA, SAR = 1, PH = 30 min	36 ± 16 (35)	37 ± 15 (35)	44 ± 13 (40)
NLA, SAR = 1, PH = 60 min	40 ± 16 (45)	44 ± 14 (50)	48 ± 12 (50)
NLA, SAR = 2, PH = 15 min	16 ± 13 (10)	16 ± 13 (10)	11 ± 7 (10)
NLA, SAR = 2, PH = 30 min	26 ± 15 (25)	27 ± 15 (25)	36 ± 15 (30)
NLA, SAR = 2, PH = 60 min	34 ± 16 (35)	34 ± 16 (35)	46 ± 13 (45)
OE, PH = 15 min	19 ± 16 (15)	18 ± 16 (10)	23 ± 19 (15)
OE, PH = 30 min	27 ± 17 (25)	26 ± 16 (25)	33 ± 22 (25)
OE, PH = 60 min	38 ± 18 (40)	37 ± 18 (40)	42 ± 18 (45)

The lower half of the table includes results of the numerical logic algorithm (NLA) and Optimal Estimation (OE) algorithm for three values of the predicted horizon (*PH*) and NLA for two values of successive alarms required (*SAR*).

^aMean ± SD values.

^bMean ± SD (median) values.

BMI, body mass index; *FPR*, false-positive rate; *TPR*, true-positive ratio; *t_w*, warning time from the first alarm to the event.

sensing reality for the subject and the scarcity of reference data, in this study CGM data were used to identify hypoglycemic events. Events were determined retrospectively and independently from the algorithm execution. The onset of a hypoglycemic event was defined to occur when CGM measurements crossed the hypoglycemia threshold and remained below for at least 10 min (i.e., three consecutive glucose measurements). At 30 min after the onset (minimum duration), the data were reassessed: recovery from the event was confirmed at the third point (not necessarily consecutive) that was at least 10 mg/dL (0.56 mmol/L) above the hypoglycemia threshold.

After the identification of each hypoglycemic event, true- and false-positive regions were defined. Algorithms presented in the literature have taken several approaches to define true-positive alarms: some used a window of ±30 min from onset¹⁰ or 45 min before onset,⁹ whereas others used a more traditional approach by declaring alarms true-positive if both the prediction and the actual value were below the hypoglycemic threshold. In this study we have defined the true positive region to be the 60-min period leading up to the hypoglycemic event to include the time period that would be considered clinically appropriate for hypoglycemia prevention or treatment. Similarly, false-positive alarms are

generally defined in the literature either as predictions that were below the threshold while the actual values were not⁹ or as a false event if not within the true-positive region.¹⁰ These metrics do not reflect the nuisance factor of false-positive alarms. Therefore, the false-positive region extended from the end of the previous event to the beginning of the true-positive region, and each alarm during this period was counted as a

false-positive. A CGM tracing from a demonstrative case study is shown in Figure 1, with dark solid circles denoting CGM measurements and the hypoglycemia threshold of 70 mg/dL is indicated by the gray dashed-dot line. The dark dashed line in the box below each tracing is the “event line” that denotes false-positive regions, the true-positive region, and hypoglycemic events. True- and false-positive alarms are

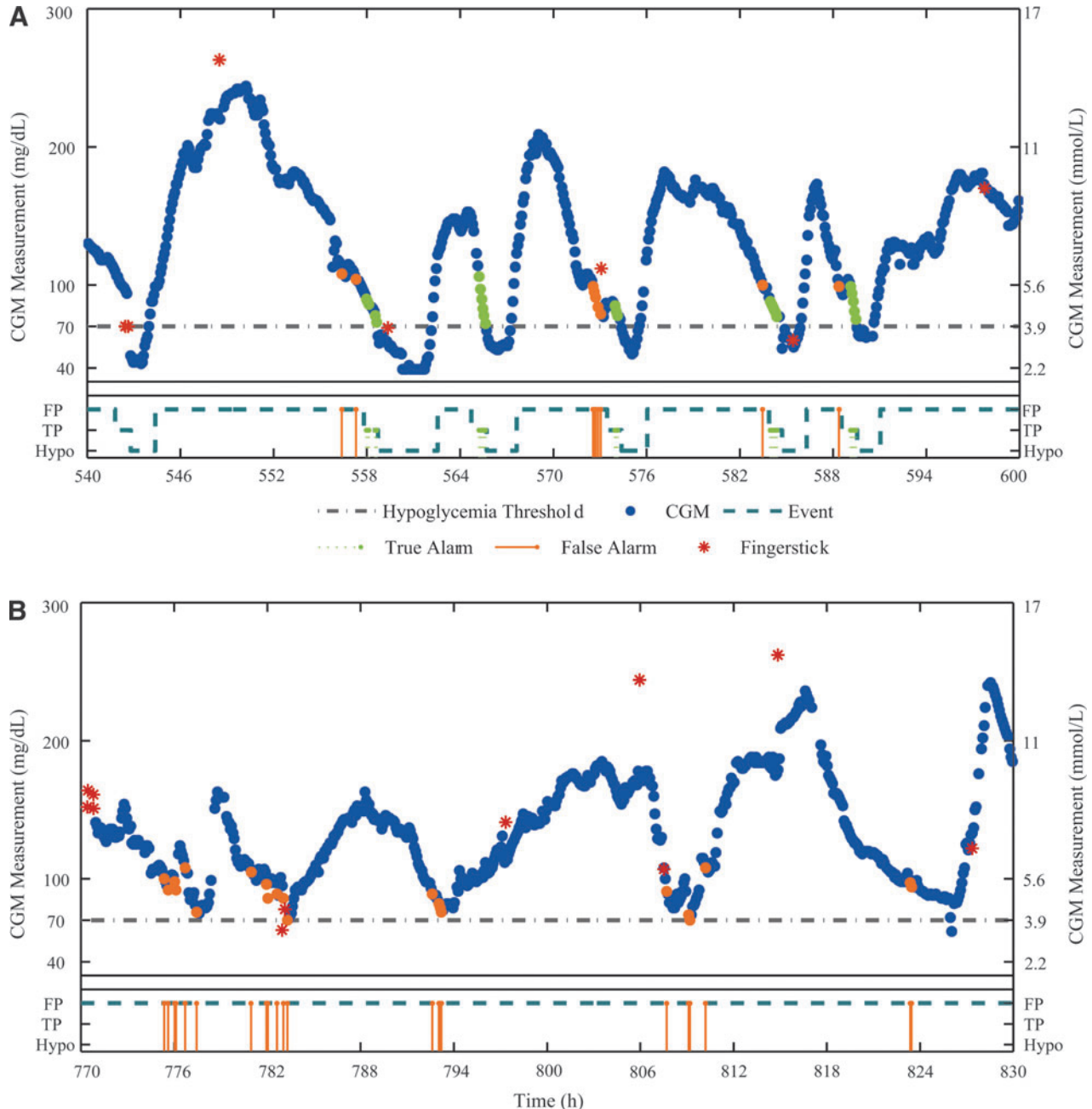


FIG. 1. Sample tracings from a demonstrative case study of ambulatory data using the Dexcom SEVEN sensor with numerical logic algorithm (NLA) with a prediction horizon of 60 min and successive alarms required of 2: (A) events occurred and (B) events did not occur. Included are 60 h of continuous glucose monitor (CGM) data (dark circles) with fingersticks measurements (asterisks) and the hypoglycemia threshold shown in gray. The event line is shown in a box below each tracing as a dark dashed line with false-positive (FP), true-positive (TP), and hypoglycemia (Hypo) regions in steps (marked on the y-axis). True alarms are in light circles, plotted coincidentally with the CGM data and on the true-positive region on the event line. False alarms are similarly plotted on the CGM data and on the false-positive region on the event line. Color images available online at www.liebertonline.com/dia

shown both as circles overlapping the CGM measurements and as stems up to the event line. Only alarms located in the true- and false-positive regions were used in the calculations below.

Metrics

The performance of the algorithms was characterized by calculating the true-positive ratio (*TPR*), the false-positive rate (*FPR*), and the warning time from the first alarm to the event (t_w). The *TPR*, commonly known as the sensitivity, is a measure of the percentage of the hypoglycemic events that were detected by the algorithm.¹⁹ In a clinical setting, the first alarm would invoke treatment or corrective action. Therefore, the first alarm in a true-positive region was considered to be a true-positive alarm, and subsequent alarms up to the event were considered associated with this alarm event and thus were not considered in the calculation of *TPR*. The *TPR* was defined as

$$TPR = \frac{TP}{P} \quad (1)$$

where *TP* is the number of true-positive alarms (alarmed hypoglycemic events) and *P* is the number of actual hypoglycemic events. Thus the *TPR* indicates the percentage of hypoglycemic events that were predicted by the algorithm.

The *FPR* is important because it quantifies the “nuisance factor” of the algorithm. A safety system that produces a large number of false-positive alarms can result in alarm fatigue or rebound hyperglycemia, depending on the method of action.²⁰ The *FPR*, expressed as false-positive alarms per day, was calculated as

$$FPR = \frac{FP}{N} \times c \quad (2)$$

where *FP* is the number of false-positive alarms, *N* is the number of alarms in the false-positive region, and *c* is the number of CGM measurements per day. Note that the *FPR* for a retrospective study such as this one may not reflect the *FPR* in a similar prospective study. In a prospective study, reactions to the alarms would have a significant impact on blood glucose trajectories. This type of study is presented in the *in silico* section.

There are several alarms in the CGM tracing in Figure 1 for the NLA algorithm. Figure 1A depicts a 60-h time period in which several events occur. The NLA parameter settings were *PH*=60 min, *SAR*=2, and the hypoglycemia threshold was 70 mg/dL (3.9 mmol/L). The resulting true alarms are shown in very light circles and false alarms by light circles, both coincident with the CGM measurement where they occurred. In addition, they are shown in the box below the tracing, both coincident with the event line. During this period, six hypoglycemic events occurred, five of which were alarmed within the true-positive region (*TPR*=83%). Nine alarms occurred within the false-positive region (*FPR*=3.6 alarms/day). Figure 1B depicts a 60-h time period in which no events occurred. In this span, 22 alarms occurred within the false-positive range (*FPR*=8.8 alarms/day). The difference between the *FPR* values in regions with several events versus periods with few events further reinforces the need for ambulatory data to evaluate algorithms, rather than specific episodes.

The warning time, t_w , between the first alarm and the beginning of a hypoglycemic episode is an important metric in algorithm comparison. It is calculated as follows:

$$t_w = t_e - t_f \quad (4)$$

where t_w is the warning time, t_e is the time of the start of the event, and t_f is the time of the first true alarm. The warning time period used in this study was 60 min: if no alarm occurred within 60 min of an event, no warning time was recorded. The warning time is an estimate of the time available for corrective action. It is common in the literature to represent the warning time as an average.^{8–10} However, when assessing a large amount of data with several hypoglycemic events, the t_w values are represented best as a distribution rather than an average to give insight into the probability of alarms occurring with a certain lead time. These results are represented in this study by a distribution plot and a precision diagram (Fig. 2). The distribution plot depicts a cumulative sum of the number of warning times in 5-min bins, plotted as a percentage of the total. The precision diagram is divided into bins of 0–15, 15–30, 30–45, and 45–60 min. The warning time for each bin was calculated as

$$r_i = \frac{w_i}{b_i} \quad (5)$$

where r_i is the warning time rate, w_i is the number of warning times, and b_i is the number of minutes, all in bin *i*. The normalized warning time rate, \tilde{r}_i , which varies from 0 to 1, was then calculated:

$$\tilde{r}_i = \frac{r_i}{r_{\max}}$$

where r_{\max} is the maximum rate in the set of bins. The normalized warning time rate is plotted in concentric circles like a dartboard, to illustrate the performance of the algorithm. For instance, if the highest warning time rate occurred within 15 min of the event, the center circle will correspond to a value of 1. The contrast with the other circles indicates whether the warning time rate is narrowly or widely distributed.

Results

Retrospective study

Results were generated for all subjects using the NLA and OE (Table 1). An example of metrics for the demonstrative case study is illustrated in Figure 2. In Figure 2A, the *TPR* values (multiplied by 100%) and *FPR* values both show an expected increase as the *PH* increases. Increasing the *SAR* resulted in a marked decrease in both *TPR* and *FPR* and a change in the shape of the *PH* trend, with a faster increase in *TPR* with *PH*.

The warning time distribution plot is in Figure 2C, which is a snapshot of the entire range of *PH* values. Because a large amount of data was analyzed, the probability that an alarm will occur within a specific time, t_w , can be deduced by this plot. As the *PH* increases, the probability that the majority of alarms would occur further before the event increases. This observation is useful when comparing parameter settings for different methods of action. For example, to prevent hypoglycemia using pump suspension, a setting that alarmed more

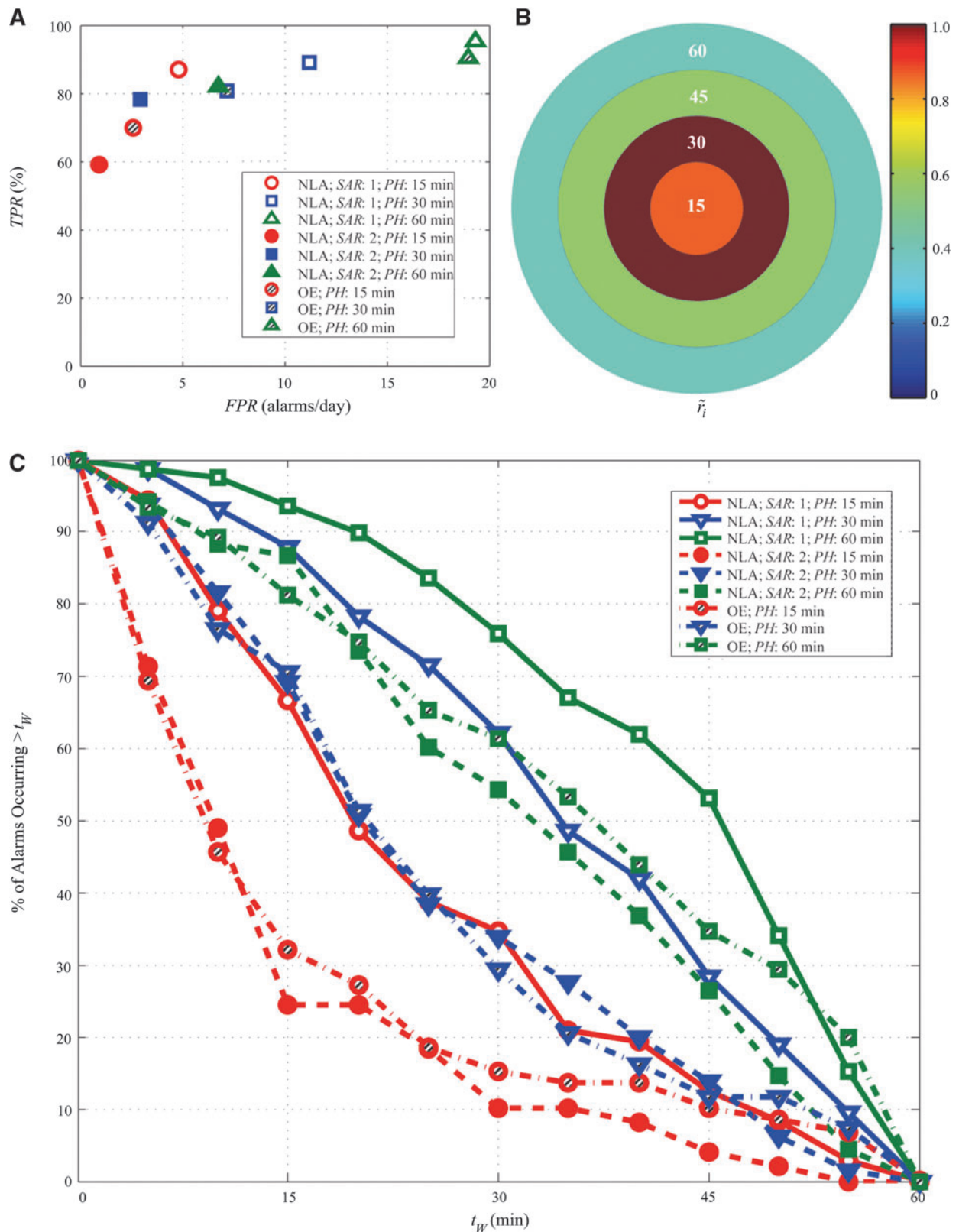


FIG. 2. Sample results from the demonstrative case study using the Dexcom SEVEN sensor. Data include 98.5 days with 83 hypoglycemia events (<70 mg/dL [3.9 mmol/L]). **(A)** True-positive ratio (TPR) values and false-positive rate (FPR) values are shown for three different prediction horizon (PH) values for the numerical logic algorithm (NLA) and the Optimal Estimation (OE) algorithm (15, 30, and 60 min in circles, squares, and triangles, respectively), with NLA with successive alarms required (SAR) of 1 (open symbols), an SAR of 2 (solid symbols), and OE (striped symbols). **(B)** The relative density of first alarms is shown for the NLA with $PH=30$ min and $SAR=2$. **(C)** The percentage of first alarms that are longer than t_W time before the event are shown for NLA and OE with three PH values (15, 30, and 60 min in circles, squares, and triangles, respectively), with an SAR of 1 (solid lines and open symbols), an SAR of 2 (dashed lines and filled symbols), and OE (dash-dotted line with striped symbols). The legend entries indicate the algorithm, the SAR, and the PH. Color images available online at www.liebertonline.com/dia

than 45 min ahead with the highest probability is required. A line drawn at 45 min would indicate that either NLA with an SAR of 1 and a PH of 60 min or OE with a PH of 60 min could be chosen, as first alarms have a probability of occurring more than 45 min ahead 53% and 35% of the time, respectively. Figure 2B displays the precision plot for the NLA with PH=30 min and SAR=2. The bin with the maximum normalized warning time rate is shown with a value of 1. This PH has moderately high precision, with the r_{\max} within Bin 2 (15–30 min), and a very similar rate in Bin 1 (87% of Bin 2) with a sharp drop-off in Bins 3 and 4 (57% and 39%, respectively, of the rate in Bin 2). This observation indicates that a large majority of the alarms would occur close to the event. With a 78% TPR and only 3.0 false alarms per day, this setting may be a good choice for an alerting application.

In silico study

The NLA was tested retrospectively to determine parameter settings with the appropriate qualities for different methods of hypoglycemia mitigation. A prospective study with induced hypoglycemia and accompanying mitigation using several parameter settings was performed in silico with the NLA, with major results in Figure 3. Each subplot displays the spectrum of CGM values for all 10 subjects for the duration of the simulation. Values below 70 mg/dL (3.9 mmol/L) and above 180 mg/dL (10 mmol/L) are shown as dark blue and red cells, respectively, to easily visualize hypo- and hyperglycemia and quickly surmise differences between protocols. One can see by the abundance of dark blue that no hypoglycemia was experienced in the control (Fig. 3A) and that severe hypoglycemia was experienced in the control with protocol (Fig. 3B). The first alarm produced from the hypoglycemia prediction algorithm is shown either as a square (false-positive alarm) or a circle (true-positive alarm) in Fig. 3A and B. For the control with protocol, 30.5 ± 5.0 units of insulin was delivered over an 18-h period, with missed meal boluses of 8.0 ± 3.8 units. There were no hypoglycemic events for the control versus 10 events for the control with protocol (14.9% of time spent below 70 mg/dL [3.9 mmol/L]).

The pump suspension protocol results are shown in Figure 3C–E for SAR=2 and PH values of 15, 30, and 60 min, with an average number of 90-min suspensions per subject of 1.1 ± 0.6 , 1.8 ± 0.4 , and 2.5 ± 0.7 , respectively. Although the duration of hypoglycemia decreased, as seen by the decrease in dark-blue cells from Fig. 3B to Fig. 3C–E, from 14.9% to 10.1%, 8.7%, and 8.3%, respectively (significant only for the 30- and 60-min PH values), the duration of hypoglycemia was significant, even with long warning times and multiple suspensions. This observation suggests that, in the event of an over-bolus or bolus without a meal, suspending the insulin pump is insufficient to avoid or alleviate subsequent hypoglycemia. Not only was hypoglycemia not avoided, but an increase in rebound hyperglycemia (over 180 mg/dL [10 mmol/L]) was observed, as seen by the introduction of red cells on the plot (difference not significant, from 0.4% to 1.9%, 3.5%, and 3.7%, respectively).

The alert protocol is shown in Figure 3F–H for SAR=2 and PH values of 15, 30, and 60 min with an average number of rescue CHO administrations per subject of 1.1 ± 0.6 , 1.6 ± 0.7 , and 1.9 ± 0.9 , respectively. It is evident that the administration of rescue CHOs significantly decreased the duration of hy-

poglycemia, displayed by very few dark-blue cells (from 14.9% to 4.7%, 0.5%, and 0.2%, respectively) and the number of episodes (from 10 to seven, two, and one, respectively) without causing rebound hyperglycemia. In addition, although the improvement of a PH of 30 min over 15 min was evident, improvement of a PH of 60 min over 30 min was slight. This reinforces the findings of the retrospective study in which the TPR of the 30- and 60-min PH values were high and similar, whereas that of the 15-min PH was much lower. This observation indicates that a low to moderate PH is sufficient to avoid hypoglycemic events when alerting, even with a large bolus without a meal.

Conclusions

The proposed metrics provide practical measures of the performance of hypoglycemia prediction algorithms in a real-world setting. They can be applied to compare and evaluate prediction algorithms using ambulatory data and to select appropriate parameter settings for the chosen method of action. For example, if the application is to use NLA to alert the subject to impending hypoglycemia and to recommend rescue CHOs, the results of the retrospective study suggest that selecting parameter settings of PH=30 min and SAR=2 may be ideal. On the other hand, for hypoglycemia prevention using pump suspension, more sensitive settings, such as PH=60 min and SAR=1, may be required. It is evident that very different parameter settings may be required depending on the method of action associated with the hypoglycemia algorithm.

The results of the retrospective study were clearly reflected in the results of the prospective study, specifically the impact of the TPR and t_w on the alert protocol. For example, with parameter settings of PH=30 min and SAR=2, the probability of alarms occurring greater than 15 or 30 min before a hypoglycemia episode was 71% and 34%, respectively (see Fig. 2C), with the greatest density of alarms occurring between 15 and 30 min (see Fig. 2B). When the subjects were treated with rescue CHOs, the parameter settings of PH=30 min and SAR=2 were able to reduce the time in hypoglycemia from 14.9% to 0.5%. All first alarms occurred at least 25 min ahead (four out of seven were 25–36 min), and eight out of 10 episodes were completely avoided. However, these parameter settings were less successful for pump suspension (8.7% of the time spent in hypoglycemia). Suspending with a setting of PH=30 min was insufficient to avoid hypoglycemia due to a large over-bolus, even with a 78% TPR and an average warning time of 27 min. With suspension, with a setting of PH=60 min was the time in hypoglycemia not greatly reduced (–8.3%). The TPR for this setting was only slightly higher (82%), with an increase in average warning time to 34 min and 54% probability of occurring more than 30 min ahead (see Fig. 2C). A much longer warning time would be necessary for treating a large over-bolus by pump suspension.

The influence of the parameter settings for the hypoglycemia prediction algorithms on its performance for a selected mitigation method such as alerting or pump suspension must be assessed before the algorithm is incorporated into an artificial pancreas system, whether current systems or a future artificial pancreas design. Application of the proposed metrics when assessing retrospective data is an easy and effective way to design a hypoglycemia prediction algorithm.

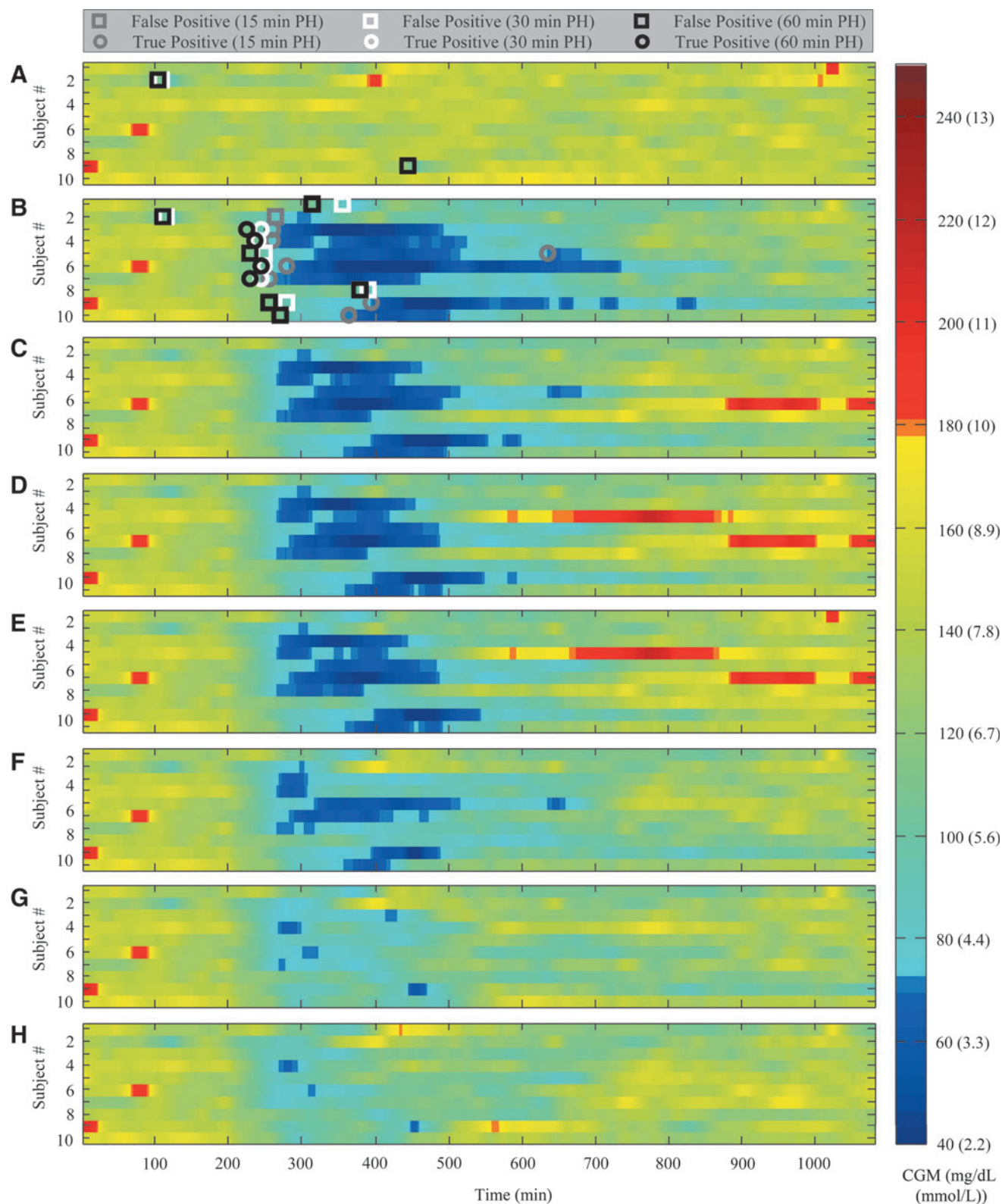


FIG. 3. Results of an in silico study on 10 adult subjects who were given their basal rate and a bolus (8.0 ± 3.8 units) for a 65-g carbohydrate meal at 150 min to simulate an over-bolus that may lead to hypoglycemia: (A) the control with no over-bolus, (B) the over-bolus protocol, (C–E) pump suspension protocol for 15-, 30-, and 60-min prediction horizon (PH) values, respectively, and (F–H) alert protocol for 15, 30 and 60 min PH values, respectively. False-positive alarms (open squares) and true-positive alarms (open circles) are shown for protocols without action (A and B). A value for successive alarms required of 2 was used throughout. CGM, continuous glucose monitor.

Acknowledgments

We acknowledge the National Institutes of Health (grants DK085628-01 and DP3DK094331-01) for their financial support. R.A.H. received financial support from the Eugene Cota-Robles Fellowship and the Air Products and Chemicals Discovery Fellowship. Product support was received from Dexcom, Inc. This research was conducted with product support from the Investigator-Initiated Study Program of LifeScan Corporation.

Author Disclosure Statement

No competing financial interests exist.

References

1. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med* 1993; 329:977–986.
2. Cryer PE, Axelrod L, Grossman AB, Heller SR, Montori VM, Seaquist ER, Service FJ; Endocrine Society: Evaluation and management of adult hypoglycemic disorders: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2009;94:709–728.
3. Workgroup on Hypoglycemia; American Diabetes Association: Defining and reporting hypoglycemia in diabetes: a report from the American Diabetes Association Workgroup on Hypoglycemia. *Diabetes Care* 2005;28:1245–1249.
4. Buckingham B, Chase HP, Dassau E, Cobry E, Clinton P, Gage V, Caswell K, Wilkinson J, Cameron F, Lee H, Bequette BW, Doyle FJ III: Prevention of nocturnal hypoglycemia using predictive alarm algorithms and insulin pump suspension. *Diabetes Care* 2010;33:1013–1017.
5. Palerm CC, Bequette BW: Hypoglycemia detection and prediction using continuous glucose monitoring—a study on hypoglycemic clamp data. *J Diabetes Sci Technol* 2007;1:624–629.
6. McGarraugh G, Bergenstal R: Detection of hypoglycemia with continuous interstitial and traditional blood glucose monitoring using the FreeStyle Navigator continuous glucose monitoring system. *Diabetes Technol Ther* 2009;11:145–150.
7. Buckingham B, Wilson DM, Lecher T, Hanas R, Kaiserman K, Cameron F: Duration of nocturnal hypoglycemia before seizures. *Diabetes Care* 2008;31:2110–2112.
8. Dassau E, Cameron F, Lee H, Bequette BW, Zisser H, Jovanovic L, Chase HP, Wilson DM, Buckingham BA, Doyle FJ III: Real-time hypoglycemia prediction suite using continuous glucose monitoring: a safety net for the artificial pancreas. *Diabetes Care* 2010;33:1249–1254.
9. Eren-Oruklu M, Cinar A, Quinn L: Hypoglycemia prediction with subject-specific recursive time-series models. *J Diabetes Sci Technol* 2010;4:25–33.
10. Cameron F, Niemeier G, Gundy-Burlet K, Buckingham B: Statistical hypoglycemia prediction. *J Diabetes Sci Technol* 2008;2:612–621.
11. Kaufman FR, Agrawal P, Lee SW, Kannard B: Characterization of the low glucose suspend feature of the Medtronic MiniMed Paradigm Veo insulin pump system and events preceding its activation [abstract]. *Diabetes* 2011;60(Suppl 1):A249.
12. Danne T, Kordonouri O, Remus K, Bläsing S, Holder M, Wadien T, Haberland H, Golembowski S, Zierow S, Hartmann R, Thomas A: The low glucose suspend (LGS) function in sensor-augmented pump therapy prevents hypoglycaemia in children [abstract 150-OR]. *Diabetes* 2011;60(Suppl 1):A41.
13. Harvey RA, Dassau E, Zisser HC, Bevier W, Seborg D, Jovanovic L, Doyle FJ III: Analysis of a hypoglycemia prediction algorithm on an extensive library of ambulatory data [abstract 884-P]. *Diabetes* 2011;60(Suppl 1):A240.
14. Garg S, Jovanovic L: Relationship of fasting and hourly blood glucose levels to HbA1c values: safety, accuracy, and improvements in glucose profiles obtained using a 7-day continuous glucose sensor. *Diabetes Care* 2006;29:2644–2649.
15. McGarraugh G: Alarm characterization for a continuous glucose monitor that replaces traditional blood glucose monitoring. *J Diabetes Sci Technol* 2010;4:49–56.
16. Kincaid D, Cheney EW: *Numerical Analysis: Mathematics of Scientific Computing*, 3rd ed. Pacific Grove, CA: Brooks/Cole, 2002.
17. Palerm CC, Willis JP, Desemone J, Bequette BW: Hypoglycemia prediction and detection using optimal estimation. *Diabetes Technol Ther* 2005;7:3–14.
18. Kovatchev BP, Breton DM, Dalla Man C, Cobelli C: In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. *J Diabetes Sci Technol* 2009;3:44–55.
19. Altman DG, Bland JM: Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994;308:1552.
20. Mastrototaro J, Welsh JB, Lee S: Practical considerations in the use of real-time continuous glucose monitoring alerts. *J Diabetes Sci Technol* 2010;4:733–739.

Address correspondence to:
 Prof. Francis J. Doyle III, Ph.D.
 Department of Chemical Engineering
 University of California, Santa Barbara
 Santa Barbara, CA 93106-5080

E-mail: frank.doyle@icb.ucsb.edu