

Assessment of Blood Glucose Predictors: The Prediction-Error Grid Analysis

Sampath Sivananthan, Ph.D.,¹ Valeriya Naumova, M.Sc.,¹ Chiara Dalla Man, Ph.D.,² Andrea Facchinetti, Ph.D.,²
Eric Renard, M.D., Ph.D.,^{3,4} Claudio Cobelli, Ph.D.,² and Sergei V. Pereverzyev, Ph.D., Dr.Sc.¹

Abstract

Background: Prediction of the future blood glucose (BG) evolution from continuous glucose monitoring (CGM) data is a promising direction in diabetes therapy management, and several glucose predictors have recently been proposed. This raises the problem of their assessment. There were attempts to use for such assessment the continuous glucose-error grid analysis (CG-EGA), originally developed for CGM devices. However, in the CG-EGA the BG rate of change is estimated from past BG readings, whereas predictors provide BG estimation ahead of time. Therefore, the original CG-EGA should be modified to assess predictors. Here we propose a new version of the CG-EGA, the Prediction-Error Grid Analysis (PRED-EGA).

Methods: The analysis is based both on simulated data and on data from clinical trials, performed in the European FP7-project "DIAdvisor." Simulated data are used to test the ability of the analyzed CG-EGA modifications to capture erroneous predictions in controlled situation. Real data are used to show the impact of the different CG-EGA versions in the evaluation of a predictor.

Results: Using the data of 10 virtual and 10 real subjects and analyzing two different predictors, we demonstrate that the straightforward application of the CG-EGA does not adequately classify the prediction performance. For example, we observed that up to 70% of 20 min ahead predictions in the hyperglycemia region that are classified by this application as erroneous are, in fact, accurate. Moreover, for predictions during hypoglycemia the assessments produced by the straightforward application of the CG-EGA are not only too pessimistic (in up to 60% of cases), but this version is not able to detect real erroneous predictions. In contrast, the proposed modification of the CG-EGA, where the rate of change is estimated on the predicted BG profile, is an adequate metric for the assessment of predictions.

Conclusions: We propose a new CG-EGA, the PRED-EGA, for the assessment of glucose predictors. The presented analysis shows that, compared with the straightforward application of the CG-EGA, the PRED-EGA gives a significant reduction of the misclassification cases. A reduction by a factor of at least 4 was observed in the study. Moreover, the PRED-EGA is much more robust against uncertainty in the input and references.

Introduction

PREDICTION OF THE FUTURE blood glucose (BG) evolution from continuous glucose monitoring (CGM) data is a promising direction in diabetes therapy management. The importance of such prediction has been shown by several applications.^{1,2} There are many possible prediction techniques,³ and several glucose predictors have been recently proposed.⁴⁻⁷ This raises the problem of their assessment.

Examples of the assessment metrics adopted in the literature include Clarke error grid analysis,⁵ mean relative abso-

lute deviation percentage,^{6,7} and the continuous glucose-error grid analysis (CG-EGA),^{7,8} with CGM data serving as the reference in all cases.

Simple illustration can demonstrate that among these metrics the CG-EGA is the most sensitive in the recognition of different prediction performances. It can be seen from Figure 1 where the horizontal line, paired with each of the other four lines, stays for a prediction versus reference plots, such that we have four performance configurations, which are, of course, different from the clinical viewpoint. For example, the prediction of a stable glucose concentration at the background

¹Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Linz, Austria.

²Department of Information Engineering, University of Padova, Padova, Italy.

³Department of Endocrinology, Diabetes, Nutrition, and CIC INSERM 1001, CHU Montpellier, Montpellier, France.

⁴CNRS FRE 3400, University of Montpellier 1, Montpellier, France.

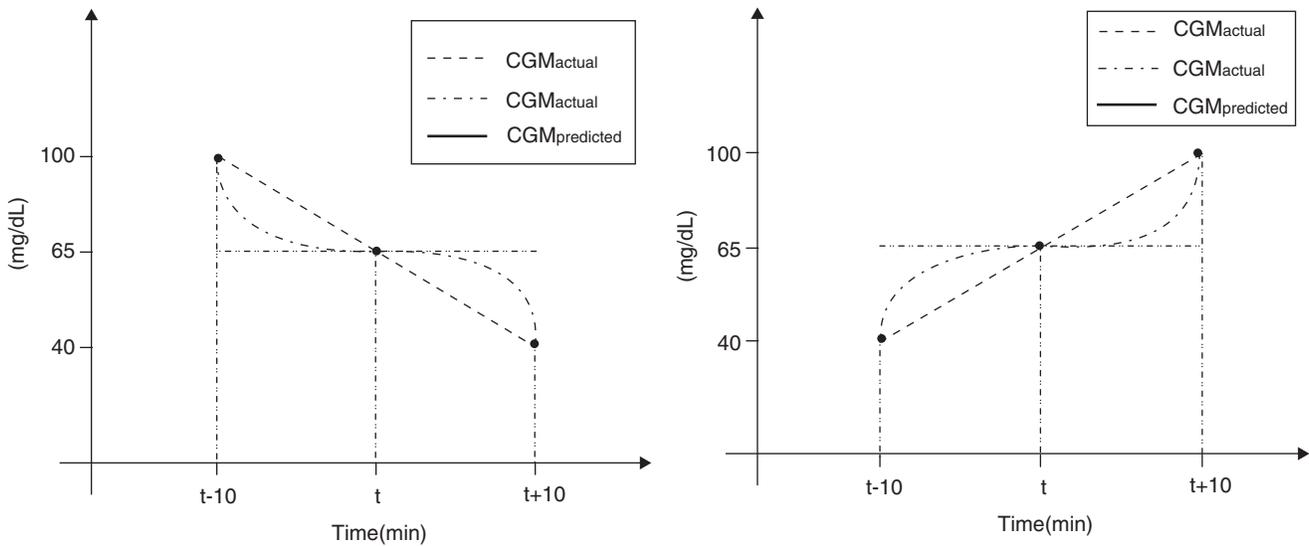


FIG. 1. Simple illustration demonstrating the sensitivity of the continuous glucose error grid analysis approach. CGM, continuous glucose monitoring.

of a fall in the hypoglycemic phase depicted in the left panel of Figure 1 should be considered more dangerous than the same prediction at the background of a recovery from the hypoglycemia shown in the right panel.

It is easy to check that for three assessment moments $t - 10$, t , and $t + 10$ (in min), traced on Figure 1, the mean relative absolute deviation percentage has the same value for all these configurations, and the Clarke error grid analysis is also not able to distinguish between them, indicating 67% in zone A and 33% in zone B for all cases.

At the same time, only the CG-EGA provides different assessment results for all performance configurations considered.

Namely, the predictions for the three moments traced in the left panel of Figure 1 are assessed as follows: 100% of erroneous predictions during hypoglycemia and 100% of benign errors during normoglycemia for the plot "horizontal line versus inclined line"; 50% of accurate and erroneous predictions during hypoglycemia and 100% of benign errors during normoglycemia for the plot "horizontal line versus curve-shaped line."

For the right panel of Figure 1 the CG-EGA yields 100% of benign errors during hypo- and normoglycemia for the plot "horizontal line versus inclined line," whereas for the plot "horizontal line versus curve-shaped line" the values are 50% of accurate and benignly erroneous predictions during hypoglycemia and 100% of benign errors during normoglycemia.

At this point it is noteworthy that demonstrated sensitivity in recognition of different prediction performances is due to the fact that the CG-EGA accounts for direction and rates of glucose change. Both of them are measured, in fact, by values of time derivatives calculated at the assessment moments for reference as well as for assessed glucose profile. Just the knowledge of these values allows us to distinguish all performance configurations in Figure 1. Therefore, only in the ideal situation of continuous signal and thus perfect derivative would the CG-EGA be the "gold standard" error grid analysis for the prediction assessment.

However, in practice the exact values of time derivatives are not given but can only be estimated by means of approximate formulae, such as, for example, a backward dif-

ference, a forward difference, or the central difference, which are well known and can be found in the Appendix.

Moreover, recall that the CG-EGA was originally developed for quantifying the accuracy of CGM devices,^{9,10} which at an assessment time, say $t = t_i$, do not provide any information for the future time step $t = t_i + \Delta t$, where Δt is a fixed elapsed time between two successive CGM readings. Therefore, in the original CG-EGA the rates of glucose change are estimated backward in time.

In contrast, predictors provide a glucose estimation ahead of time, and it paves a new way for estimating the rates of glucose change. As a result, several versions of the CG-EGA can be proposed for the prediction assessment.

The aim of this study is to analyze them with respect to adequateness to judge upon the prediction accuracy and to robustness to approximation and data errors. By resorting to the analysis of the performance of a particular glucose predictor on several simulated and clinical data sets, this work shows that the straightforward application of the original CG-EGA, as it was developed for CGM systems, does not adequately classify the prediction performance. In contrast, a proposed modification of the CG-EGA, where the rates of change are estimated on predicted and reference profiles by means of the central difference, is a very rigorous metric for the prediction assessment.

We also argue that the conclusion that the proposed modification of the CG-EGA is more adequate in assessing glucose predictors than the original CG-EGA, is general, and does not depend on a particular predictor. To demonstrate this, we apply the compared metrics to assess the performance of the simplest glucose predictor known as "projection ahead" of the current glucose trend.¹¹ Then the assessment results support the conclusion made on the basis of the predictor that was used first.

Experimental Procedures

Versions of the CG-EGA for glucose predictors

Following the CG-EGA methodology^{9,10} to assess the accuracy of a glucose predictor, one should record reference

glucose estimates paired with the estimates predicted for the same moments and look at the two essential aspects of the clinical accuracy: rate error grid analysis (R-EGA) and point error grid analysis (P-EGA). The modifications of the CG-EGA we consider are based on the same rate error grid and point error grid as proposed for the assessment of CGM systems.^{9,10} The difference between these modifications is in the way of estimating the rate of glucose change.

Let $\{X_{\text{reference}}(t_i), i = 1, 2, \dots\}$ be glucose estimates that are used as references in the prediction assessment. Using the well-known approximate formulae presented in the Appendix and assuming that the reference estimates are taken every $\Delta t = t_i - t_{i-1}$ min, one can estimate the reference rate of change in one of the following ways:

$$X'_{\text{reference}}(t_i) \approx \frac{X_{\text{reference}}(t_i) - X_{\text{reference}}(t_{i-1})}{\Delta t} \quad (1)$$

$$X'_{\text{reference}}(t_i) \approx \frac{X_{\text{reference}}(t_{i+1}) - X_{\text{reference}}(t_i)}{\Delta t} \quad (2)$$

$$X'_{\text{reference}}(t_i) \approx \frac{X_{\text{reference}}(t_{i+1}) - X_{\text{reference}}(t_{i-1})}{2\Delta t} \quad (3)$$

Let now $X(t, \tau)$ be a predicted value of the glucose concentration at time τ (min) in the future, when the prediction is made at time $t < \tau$. Then for a fixed prediction horizon PH a predicted value of the glucose concentration $X_{\text{predicted}}(t_i)$ for the future time $t = t_i$ is given as $X_{\text{predicted}}(t_i) = X(t_i - PH, t_i)$.

From a time series viewpoint each item of the reference time series $\{X_{\text{reference}}(t_i), i = 1, 2, \dots\}$ should be compared with the corresponding item of the predicted time series $\{X_{\text{predicted}}(t_i), i = 1, 2, \dots\}$, and in a straightforward application of the original CG-EGA to the prediction assessment the predicted rate of change is estimated backward in time as follows:

$$\begin{aligned} X'_{\text{predicted}}(t_i) &\approx \frac{X_{\text{predicted}}(t_i) - X_{\text{predicted}}(t_{i-1})}{\Delta t} \\ &= \frac{X(t_i - PH, t_i) - X(t_{i-1} - PH, t_{i-1})}{\Delta t} \end{aligned} \quad (4)$$

and then plotted against the reference rate given by Eq. 1.

This version of the CG-EGA has been already used in the literature for the assessment of the glucose predictors,^{7,8} and we will refer to it as the Prediction-Error Grid Analysis (PRED-EGA), version 0.

The other versions of the PRED-EGA, which we are going to present, originate from the observation that for a fixed PH all the predictors described in the literature³⁻⁷ can predict not just the glucose concentration at time $t = t_i$, but the whole future glucose profile as a function of time $X_{\text{predicted}}(t) = X(t_i - PH, t)$ for $t \geq t_i - PH$. Then the predicted rate of change $X'_{\text{predicted}}$ at $t = t_i$ can be estimated in accordance with the well-known formulae as follows:

$$X'_{\text{predicted}}(t_i) \approx \frac{X(t_i - PH, t_i) - X(t_i - PH, t_{i-1})}{\Delta t} \quad (5)$$

$$X'_{\text{predicted}}(t_i) \approx \frac{X(t_i - PH, t_{i+1}) - X(t_i - PH, t_i)}{\Delta t} \quad (6)$$

$$X'_{\text{predicted}}(t_i) \approx \frac{X(t_i - PH, t_{i+1}) - X(t_i - PH, t_{i-1})}{2\Delta t} \quad (7)$$

Thus, the three well-known approximate formulae for the first derivative lead to three different versions of the PRED-EGA, where the reference rate of change and its predicted coun-

terpart are estimated by the pairs of formulae Eqs. 1 and 5, Eqs. 2 and 6, or Eqs. 3 and 7. In the following we will refer to them as the PRED-EGA, versions 1, 2, and 3, respectively.

As we already mentioned, in practice CGM readings, which are used for predictions and for references, are given with a fixed elapsed time Δt . For several widely used CGM sensors we have $\Delta t = 5$ (min). It means that in above-mentioned formulae for rates of change the value of Δt cannot be chosen at will. Therefore, the reliability and robustness of these approximate formulae with fixed Δt should be tested prior to using them on clinical data.

Such reliability and robustness tests can be made with simulated data, which allow an accurate calculations of $X'_{\text{reference}}(t)$ and $X'_{\text{predicted}}(t)$ at the assessment points $t = t_i$, $i = 1, 2, \dots$. In this controlled situation the accurate values of $X'_{\text{reference}}(t_i)$ and $X'_{\text{predicted}}(t_i)$ can be used in the R-EGA that will lead to the ideal version of the PRED-EGA, which has been already characterized as the "gold standard" error grid analysis for the prediction assessment. Comparing the assessments provided by this "gold standard" metric with ones given by the versions of the PRED-EGA described above, we can determine the most adequate of them. Below we present the results of such comparison.

Subjects and dataset

The dataset contains the data of 100 virtual subjects and 10 real subjects. The virtual subjects data are obtained from Padova/University of Virginia simulator.¹² For each in silico patient the CGM readings have been simulated and sampled with a frequency of 1 min during 3 days. We perform our reliability and robustness tests with data of 10 virtual subjects provided by the Padova/University of Virginia simulator.¹² Data of the first three subjects (Virtual ID 1, 2, and 3) have been randomly chosen from 100 traces generated by the simulator. The other seven data sets (Virtual ID 17, 18, 24, 33, 34, 42, and 47) have been chosen because they contain both hypo- and hyperglycemic periods.

The data of 10 real subjects were collected during two clinical trials (1A and 1B) that were performed in the Centre d'Investigation Clinique du CHU de Montpellier, Montpellier, France within the framework of the European FP7-project "DIAdvisor." The subjects considered have been diagnosed with type 1 diabetes and treated with insulin for at least 12 months before the trials; their ages are between 18 and 70 years. During the study, the BG estimates were sampled every 5 min using a CGM sensor. In the "DIAdvisor 1B" trial such information was collected during 3 days, whereas in the "DIAdvisor 1A" trial it was done during 2 days. The patients consumed three meals per day (40 g of carbohydrates at breakfast, 70 g at lunch and dinner) and decided on their insulin need according to the current BG level.

Glucose predictors used in the study

It is clear that a comparison of different versions of the PRED-EGA can only be done on the basis of some glucose predictor. In our study the role of such a basic predictor is played by an algorithm, which is based on adaptive kernel methods.^{13,14}

The current version of the algorithm is portable from individual to individual without being readjusted. It was trained on the data, which are not used in the present study.

TABLE 1. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE "GOLD STANDARD" PREDICTION-ERROR GRID ANALYSIS FOR THE BASIC PREDICTOR OPERATING ON SIMULATED NOISE-FREE DATA WITH A PREDICTION HORIZON OF 20 MIN

Virtual ID	BG ≤ 70 mg/dL			BG 70–180 mg/dL			BG ≥ 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
1	—	—	—	100	—	—	100	—	—
2	—	—	—	100	—	—	100	—	—
3	—	—	—	100	—	—	—	—	—
17	100	—	—	100	—	—	—	—	—
18	100	—	—	100	—	—	—	—	—
24	100	—	—	100	—	—	—	—	—
33	100	—	—	94.84	5.15	—	100	—	—
34	100	—	—	96.98	3.02	—	100	—	—
42	100	—	—	100	—	—	100	—	—
47	100	—	—	96.67	3.33	—	100	—	—

BG, blood glucose.

The performance of the current version of the algorithm has been demonstrated recently in the presentation of the project "DIAdvisor."¹⁵ An input for the algorithm consists of six CGM readings from the past 25 min. Note that in contrast to some other studies,^{5,7,8} no data preprocessing/smoothing has been made before applying our basic glucose predictor.

Moreover, to demonstrate that the conclusions of this study are related not only with our basic predictor, we also analyze in the Appendix the performance of the simplest glucose predictor, which is based on the linear extrapolation and uses the current and 5-min-old CGM readings as the prediction input.

Results

Tests with simulated data against the "gold standard" PRED-EGA

As we already mentioned, to implement the ideal version of the PRED-EGA, which will be used as the "gold standard" assessment metric, one should know accurate values of the derivatives $X_{\text{reference}}'(t)$ and $X_{\text{predicted}}'(t)$ at the assessment points $t = t_i$, $t = t_i$, $i = 1, 2, \dots$

Such knowledge is available when one deals with simulated data used as predictor inputs and references. Table 1 demonstrates the performance assessment matrix given by the ideal version of the PRED-EGA for 20 min ahead glucose predictions, when our basic predictor is applied to the set of

simulated data. The assessment matrix is presented in the format that is usual for CG-EGA methodology.⁹

Note that even through the simulator can produce virtual CGM readings every 1 min, in our tests we try to mimic the data flow from some widely used CGM sensors, when a new reading appears every 5 min and updates the prediction input. It means that a new predicted glucose profile is produced every $\Delta t = 5$ min, and all versions of the PRED-EGA (including the "gold standard") are applied to assess the predictions at time moments $t_i = i \times \Delta t$, $i = 1, 2, \dots$ Moreover, in the PRED-EGA, versions 0–3, the formulae Eqs. 1–7 are used with $\Delta t = 5$ (min).

In Tables 2 and 3 we present the performance assessment matrices given by the PRED-EGA, versions 0 and 3, respectively (for the sake of brevity we do not present here the matrices given by other versions).

To quantify the differences in the assessment between the "gold standard" and other versions of the PRED-EGA, we use one of usual matrix norms that measures the distance between two (assessment) matrices $\{a_{ij}\}$ and $\{b_{ij}\}$, $i = 1, 2, \dots, 10$, $j = 1, 2, \dots, 9$, as follows;

$$\|\{a_{ij}\} - \{b_{ij}\}\| = \max \left\{ \sum_{j=1}^9 |a_{ij} - b_{ij}|, i = 1, 2, \dots, 10 \right\}$$

Here a_{ij} or b_{ij} is a percentage of accurate, benign, or erroneous predictions indicated in the i -th row and j -th column of the

TABLE 2. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 0, FOR THE BASIC PREDICTOR OPERATING ON SIMULATED NOISE-FREE DATA WITH A PREDICTION HORIZON OF 20 MIN

Virtual ID	BG ≤ 70 mg/dL			BG 70–180 mg/dL			BG ≥ 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
1	—	—	—	99.28	—	0.72	100	—	—
2	—	—	—	100	—	—	100	—	—
3	—	—	—	100	—	—	—	—	—
17	100	—	—	100	—	—	—	—	—
18	100	—	—	100	—	—	—	—	—
24	100	—	—	100	—	—	—	—	—
33	100	—	—	99.4	—	0.6	100	—	—
34	100	—	—	97.48	2.52	—	57.14	42.86	—
42	100	—	—	97.16	2.34	0.5	75	25	—
47	100	—	—	96.03	3.34	0.63	100	—	—

BG, blood glucose.

TABLE 3. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 3, FOR THE BASIC PREDICTOR OPERATING ON SIMULATED NOISE-FREE DATA WITH A PREDICTION HORIZON OF 20 MIN

Virtual ID	BG ≤ 70 mg/dL			BG 70–180 mg/dL			BG ≥ 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
1	—	—	—	100	—	—	100	—	—
2	—	—	—	100	—	—	100	—	—
3	—	—	—	100	—	—	—	—	—
17	100	—	—	100	—	—	—	—	—
18	100	—	—	100	—	—	—	—	—
24	100	—	—	100	—	—	—	—	—
33	100	—	—	98.41	1.59	—	100	—	—
34	100	—	—	98.15	1.85	—	100	—	—
42	100	—	—	100	—	—	100	—	—
47	100	—	—	98.33	1.67	—	100	—	—

BG, blood glucose.

corresponding assessment matrix (for example, for the matrix in Table 3, $a_{55}=0$).

In the present context the meaning of such distance is that for each particular subject the differences in the prediction assessment are summed up over glycemic regions, and then for a group of subjects considered the largest such quantity is taken as the distance between assessment matrices. It is clear that the larger such distance is, the more difference in the assessment is.

In the first row of Table 4 one can find the distances between the assessment matrices, produced by different versions of the PRED-EGA, and the matrix appearing as a result of the assessment by means of the “gold standard” PRED-EGA. From these distance values one can conclude that versions 2 and 3 are much closer to the “gold standard” than version 0.

The first row of Table 4 is related with the situation, where a virtual CGM device (the simulator) is assumed to provide accurate (reference) glucose readings. A more realistic situation can be modeled by adding white noise to the simulated readings and then by using them as predictor inputs and references. In this way we can assess the effect of the noise on the choice of the best version of the metric.

The prediction assessment matrices, produced in such a situation by versions 0, 2, and 3 of the PRED-EGA, have been calculated for the case when simulated CGM readings have been corrupted by white noise with an SD of 6 (mg/dL). The distances between these matrices and the assessment matrix from Table 1 produced by the “gold standard” PRED-EGA are presented in the second row of Table 4.

TABLE 4. DIFFERENCE IN THE ASSESSMENT BETWEEN THE “GOLD STANDARD” AND OTHER VERSIONS OF THE PREDICTION-ERROR GRID ANALYSIS MEASURED IN TERMS OF DISTANCES BETWEEN PERFORMANCE ASSESSMENT MATRICES

	Version 0	Version 2	Version 3
Simulated noise-free data	86.72	6.35	7.13
Simulated noisy data	118.96	25.34	7.11

From Table 4 it follows that among the versions considered, version 3 of the PRED-EGA is the most reliable and robust because it stays closer to the “gold standard” for both noise-free and noisy simulated data.

Assessment of the prediction from clinical data

In this section we compare the assessments provided by the PRED-EGA version 3 for the performance of our basic predictor on data from two clinical trials with those that are provided by the PRED-EGA version 0, and demonstrate that the first ones are more adequate.

As reference and predictor inputs we use the BG estimates provided every 5 min by CGM systems for five patients involved in the “DIAdvisor 1A” trial (their ID start with the letter A; the data were collected during 2 days) and for the same number of patients taking part in the “DIAdvisor 1B” trial (ID numbers start with the letter B; the data were collected during 3 days).

Prediction assessments given by the above-mentioned versions of the PRED-EGA are presented in Tables 5 and 6.

Using the matrix distance defined above, we can observe essential differences in assessments produced by the versions of the PRED-EGA under study. For matrices from the Tables 5 and 6 this distance is equal to 256.81.

However, dealing with real clinical data one is not able to implement the ideal version of the PRED-EGA and use it as the “gold standard” for a comparison with other versions and for making a conclusion about their adequateness.

At the same time, from extensive case analysis it turns out that up to 70% of predictions in the range of hyperglycemia, which are classified by the PRED-EGA version 0 as erroneous, are, in fact, accurate, and they are properly classified as such by the PRED-EGA version 3.

To illustrate this, let us closely look at the prediction assessments for two representative subjects A 121 and A 131, for which the difference between versions 0 and 3 is the most visible.

From the first two rows of Table 7 one can see that for the two patients considered version 0 failed in detection of erroneous predictions, whereas version 3 really detected them.

The situation is similar for the prediction assessment during normo- and hypoglycemia.

TABLE 5. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 0, FOR THE BASIC PREDICTOR OPERATING ON CLINICAL DATA WITH A PREDICTION HORIZON OF 20 MIN

ID	BG \leq 70 mg/dL			BG 70–180 mg/dL			BG \geq 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
A 103	73.33	16.67	10	77.16	16.98	5.86	65.36	22.88	11.76
A 108	81.82	9.09	9.09	82.97	13.32	3.71	69.57	24.64	5.8
A 118	71.43	14.29	14.29	81.21	14.24	4.55	77.38	14.29	8.33
A 121	74.55	10.91	14.55	77.6	14.2	8.2	63.16	24.81	12.03
A 131	100	—	—	61.63	27.91	10.47	70.73	17.42	11.85
B 112	66.67	15.15	18.18	67.97	22.55	9.48	48.56	31.69	19.75
B 125	30.77	23.08	46.15	56.42	33.49	10.09	59.35	26.71	13.95
B 130	77.27	11.36	11.36	72.35	21.84	5.80	82.69	12.5	4.81
B 133	100	—	—	77.32	17.53	5.15	75.24	16.93	7.84
B 136	80	6.67	13.33	70.36	21.91	7.73	67.2	23.81	8.99

BG, blood glucose.

As an example, we can present the assessments for another two patients for whom the difference between versions 0 and 3 is again the most visible.

From the last two rows of Table 7 one can see that version 0 either misclassifies predictions as erroneous (subject B 130) or is not able to detect really erroneous ones (subject B 136). Typical examples are given in Figures 2 and 3.

One may think that the conclusion about the inadequateness of the PRED-EGA version 0 for the assessment of predictions is only true with respect to the basic predictor used in the above study. To show that it is not the case, we apply both versions 0 and 3 of the PRED-EGA to assess the performance of the simplest glucose predictor, which is based on the linear extrapolation of the current and past CGM values. Such a predictor is also known as “projection ahead” of the current glucose trend.¹¹

In Tables A1 and A2 of the Appendix we present the performance assessment matrices (the distance between them is equal to 299.63) given by the PRED-EGA versions 0 and 3 for this predictor, when it operates on the same clinical data as our basic predictor. These tables allow the same conclusion as the analysis reported above for the basic predictor: namely, for predictions during normo- and hyperglycemia, the assessment given by the PRED-EGA version 0 is too pessimis-

tic—in the majority of cases (up to 70%) the predictions, which are classified as erroneous or benignly erroneous, are, in fact, accurate and properly classified in the PRED-EGA version 3. Table A3 of the Appendix presents an example of this situation for a representative subject (A 121).

Moreover, for predictions during hypoglycemia the assessment given by version 0 is not only too pessimistic, but this version of the PRED-EGA is not able to detect real errors. An example for a representative subject (A 103) is presented in Table A3 of the Appendix.

In conclusion, independently of the predictor used, PRED-EGA version 3 is the most accurate metric to assess predictor performance.

Conclusions

The main message of this study is that the straightforward application of the CG-EGA methodology in the form developed for quantifying the accuracy of CGM devices does not adequately classify the performance of glucose predictors.

At the same time, several prediction-oriented modifications of the CG-EGA can be proposed. In this study we have analyzed them and suggested the most reliable and robust one. We call this new version of the CG-EGA the “Prediction-

TABLE 6. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 3, FOR THE BASIC PREDICTOR OPERATING ON CLINICAL DATA WITH A PREDICTION HORIZON OF 20 MIN

ID	BG \leq 70 mg/dL			BG 70–180 mg/dL			BG \geq 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
A 103	80	6.67	13.33	89.23	10.46	0.31	88.89	4.58	6.54
A 108	81.82	9.09	9.09	92.37	6.97	0.65	89.86	5.8	4.35
A 118	71.43	—	28.57	96.68	3.32	—	91.67	2.98	5.36
A 121	83.64	7.27	9.09	88.99	7.23	3.77	92.48	6.02	1.5
A 131	88.89	11.11	—	90.17	9.83	—	93.38	3.14	3.48
B 112	84.85	3.03	12.12	87.58	9.48	2.94	79.1	8.61	12.3
B 125	38.46	7.69	53.85	84.02	14.16	1.83	85.46	6.53	8.01
B 130	95.45	—	4.55	94.22	4.76	1.02	93.27	0.96	5.77
B 133	—	100	—	89.74	7.69	2.57	91.22	3.14	5.64
B 136	73.33	—	26.67	94.86	5.14	0	86.77	7.41	5.82

BG, blood glucose.

TABLE 7. QUANTIFICATION OF ERRONEOUS PREDICTIONS FOR REPRESENTATIVE SUBJECTS

ID	Number of predictions that are classified as erroneous		Number of them that are misclassified as erroneous	
	Version 3	Version 0	Version 3	Version 0
A 121	2	16	0	12
A 131	10	34	0	24
B 130	2	5	0	3
B 136	4	2	0	0

Error Grid Analysis" (PRED-EGA, version 3) and recommend it as one of the assessment tools of choice when measuring the accuracy of glucose predictors.

At the same time, the goal of this article is also to attract attention to the problem of the assessment of glucose predictors. In particular, another modifications of the CG-EGA can be studied, where a dynamic zone adjustment in the point-error grid is governed by the values of PHs.

Another direction of research is the study of a new index that has been introduced recently for comparing CGM glucose prediction algorithms.¹⁶

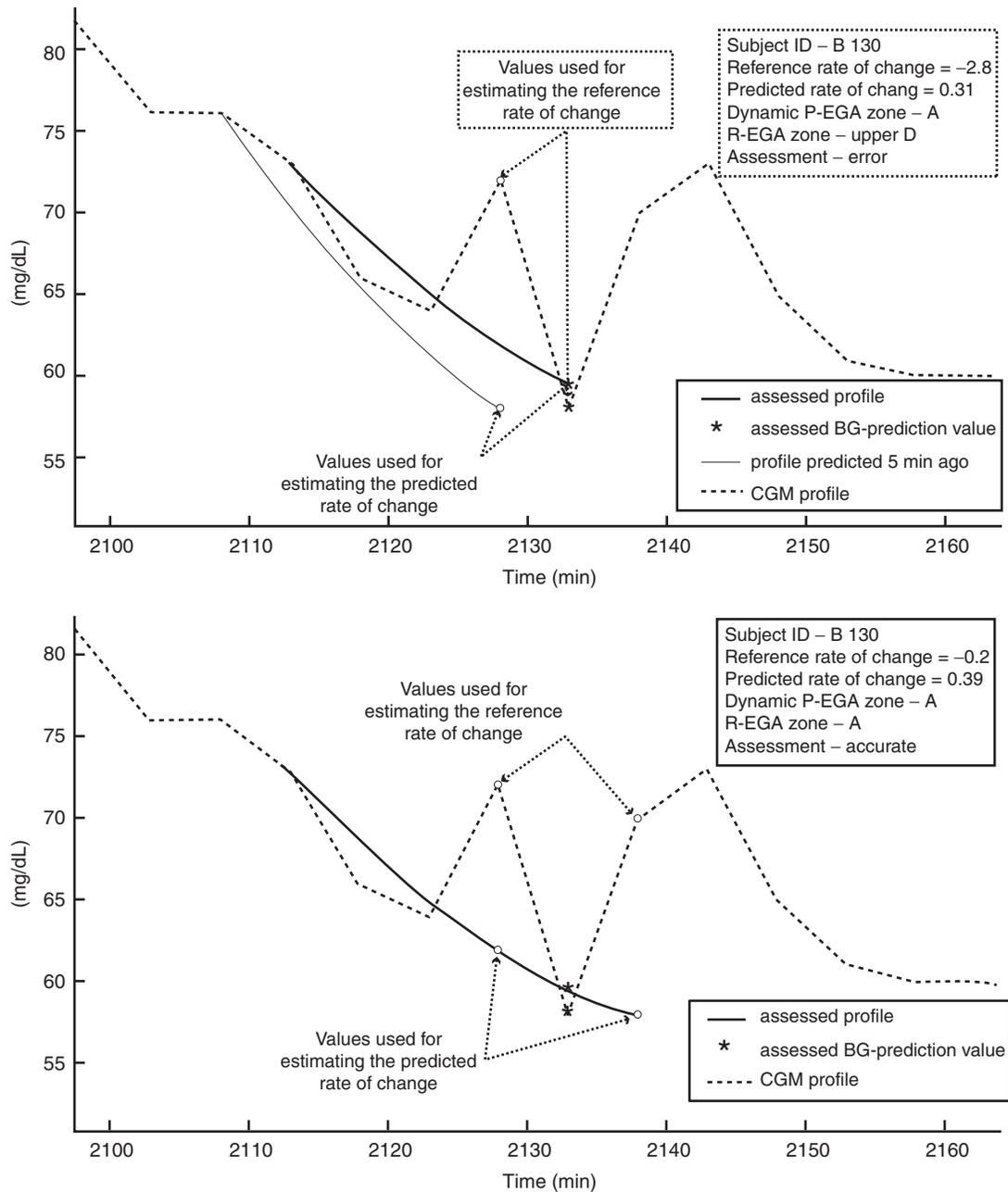


FIG. 2. A representative example of the accurate prediction during hypoglycemia, which is (bottom panel) properly classified by the Prediction-Error Grid Analysis version 3 and (top panel) misclassified by the Prediction-Error Grid Analysis version 0. BG, blood glucose; CGM, continuous glucose monitoring; P- and R-EGA, point and rate error grid analysis, respectively.

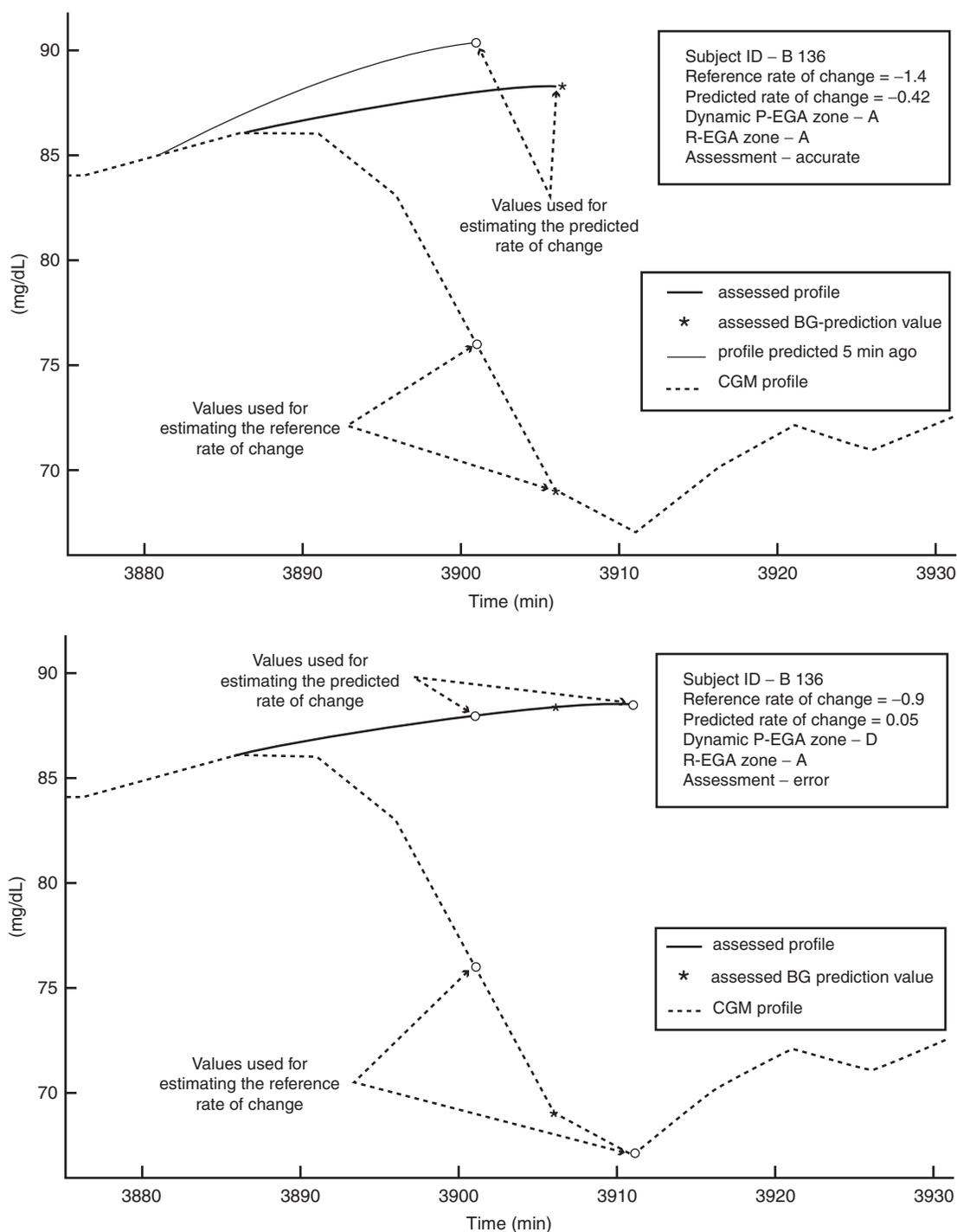


FIG. 3. A representative example of the erroneous prediction during hypoglycemia, which is (**bottom panel**) properly classified by the Prediction-Error Grid Analysis version 3 and (**top panel**) misclassified by the Prediction-Error Grid Analysis version 0. BG, blood glucose; CGM, continuous glucose monitoring; P- and R-EGA, point and rate error grid analysis, respectively.

Acknowledgments

This study has been pursued because of the need for the evaluation of glucose predictors in the European FP7-project "DIAdvisor" (www.diadvisor.eu/).

Author Disclosure Statement

The authors state that there are no competing financial interests that are related to this article.

References

1. 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.
2. 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;33:1013–1018.

3. Bequette BW: Continuous glucose monitoring: real-time algorithms for calibration, filtering and alarms. *J Diabetes Sci Technol* 2010;4:404–418.
4. Sparacino G, Zanderigo F, Corazza S, Maran A: Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Trans Biomed Eng* 2007;54:931–937.
5. Reifman J, Rajaraman S, Gribok A, Kenneth Ward W: Predictive monitoring for improved management of glucose levels. *J Diabetes Sci Technol* 2007;1:478–486.
6. Pappada SM, Cameron BD, Rosman PM: Development of neural network for prediction of glucose concentration in Type 1 diabetes patients. *J Diabetes Sci Technol* 2008;2:792–801.
7. Eren M, Cinar A, Quinn L, Smith D: Estimation of future glucose concentrations with subject-specific recursive linear models. *Diabetes Technol Ther* 2009;11:243–253.
8. Zanderigo F, Sparacino G, Kovatchev B, Cobelli C: Glucose prediction algorithms from continuous monitoring data: assessment of accuracy via continuous glucose error-grid analysis. *J Diabetes Sci Technol* 2007;1:645–651.
9. Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL: Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose-error grid analysis illustrated by TheraSense Freestyle Navigator. *Diabetes Care* 2004;27:1922–1928.
10. Clarke WL: The original Clarke error grid analysis (EGA). *Diabetes Technol Ther* 2005;7:776–779.
11. 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.
12. Kovatchev BP, Breton MD, 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.
13. De Vito E, Pereverzev SV, Rosasco L: Adaptive kernel methods using the balancing principle. *Found Comput Math* 2010;10:455–479.
14. Pereverzev SV, Sivananthan S: Regularized learning algorithm for prediction of blood glucose concentration in “No Action Period.” In: Nithiarasu P, Lohner R, Van Loon R, eds. In: 1st International Conference on Mathematical and Computational Biomedical Engineering. Swansea, UK: CMBE, 2009:395–398.
15. Poulsen JU, Avogaro A, Chaudhary F, Cobelli C, Johansson R, Nita L, Pogose M, del Re L, Renard E, Sivananthan S, Saudek F, Skillen M, Soendergaard J: A diabetes management system empowering patients to reach optimised glucose control: from monitor to advisor. In: 32nd Annual International IEEE EMBS Conference. Buenos Aires: EMBC, 2010, Paper FrE14.3.
16. Facchinetti A, Sparacino G, Trifoglio E, Cobelli C: A new index to optimally design and compare CGM glucose prediction algorithms. *Diabetes Technol Ther* 2011;13:111–119.
17. Lu S, Pereverzev SV: Numerical differentiation from a viewpoint of regularization theory. *Math Comput* 2006;75:1853–1870.
18. Mathé P, Pereverzev SV: The use of higher order finite difference schemes is not dangerous. *J Complexity* 2009;25:3–10.

Address correspondence to:
Prof. Sergei V. Pereverzyev, Ph.D., Dr.Sc.
Johann Radon Institute for Computational
and Applied Mathematics
Austrian Academy of Sciences
Altenbergerstrasse 69
Linz, 4040, Austria

E-mail: sergei.pereverzyev@oeaw.ac.at

Appendix (Tables A1–A3)

Approximate formulae for the rate of change of f(t) (first derivative)

A backward difference:

$$f'(t) \approx \frac{f(t) - f(t - h)}{h} \tag{A1}$$

A forward difference:

$$f'(t) \approx \frac{f(t + h) - f(t)}{h} \tag{A2}$$

The central difference:

$$f'(t) \approx \frac{f(t + h) - f(t - h)}{2h} \tag{A3}$$

where t is the assessment moment, h is the spacing, and f is a function of time, which describes either a reference or an assessed profile. Note that a proper choice of the spacing h can make the formulae Eqs. A1–A3 robust against data noise,^{17,18} but because in practice a reference profile is given as a discrete time series $\{f(t_i), i = 1, 2, \dots\}$ with a fixed elapsed time $\Delta t = t_i - t_{i-1}$, one usually is forced to choose $h = \Delta t$.

TABLE A1. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 0, FOR THE LINEAR EXTRAPOLATION PREDICTOR OPERATING ON CLINICAL DATA WITH A PREDICTION HORIZON OF 20 MIN

ID	BG \leq 70 mg/dL			BG 70–180 mg/dL			BG \geq 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
A 103	60	33.33	6.67	61.85	29.85	8.31	53.59	28.1	18.3
A 108	36.36	45.45	18.18	60.35	33.55	6.1	59.42	27.54	13.04
A 118	57.14	42.9	—	50.45	42.9	6.65	52.98	35.71	11.31
A 121	49.12	38.6	12.28	61.9	28.57	9.52	46.62	39.85	13.53
A 131	33.33	55.56	11.11	46.82	37.57	15.61	56.45	32.06	11.5
B 112	60.61	21.21	18.18	49.35	39.87	10.78	39.75	38.93	21.31
B 125	23.08	38.46	38.46	40.99	44.59	14.41	44.91	38.32	16.77
B 130	45.45	36.36	18.18	51.02	40.14	8.84	63.46	28.85	7.69
B 133	100	—	—	64.65	28.79	6.57	56.33	32.91	10.76
B 136	33.33	33.33	33.33	44.47	46.27	9.25	48.15	41.8	10.05

BG, blood glucose.

TABLE A2. THE PERFORMANCE ASSESSMENT MATRIX GIVEN BY THE PREDICTION-ERROR GRID ANALYSIS, VERSION 3, FOR THE LINEAR EXTRAPOLATION PREDICTOR OPERATING ON CLINICAL DATA WITH A PREDICTION HORIZON OF 20 MIN

ID	BG \leq 70 mg/dL			BG 70–180 mg/dL			BG \geq 180 mg/dL		
	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)	Accurate (%)	Benign (%)	Error (%)
A 103	80	10	10	84.66	10.74	4.6	76.47	13.07	10.46
A 108	54.55	27.27	18.18	88.7	9.35	1.96	71.01	21.74	7.25
A 118	85.71	14.29	—	92.17	4.82	3.01	88.1	5.36	6.55
A 121	68.97	22.41	8.62	84.44	9.52	6.03	77.44	18.05	4.51
A 131	77.78	11.11	11.11	79.89	13.79	6.32	87.11	6.97	5.92
B 112	75.76	9.09	15.15	82.03	12.42	5.56	66.12	14.29	19.59
B 125	23.08	38.46	38.46	70.4	20.63	8.97	77.84	11.98	10.18
B 130	79.55	4.55	15.91	83.05	13.56	3.39	87.5	6.73	5.77
B 133	—	100	—	86.93	9.55	3.52	83.86	10.44	5.7
B 136	53.33	6.67	40	88.21	8.72	3.08	80.42	14.29	5.29

BG, blood glucose.

TABLE A3. QUANTIFICATION OF ERRONEOUS PREDICTIONS GIVEN BY THE LINEAR EXTRAPOLATION PREDICTOR FOR REPRESENTATIVE SUBJECTS

ID	As erroneous		As benign	
	Version 3	Version 0	Version 3	Version 0
A 121				
Number of predictions that are classified	6	18	24	53
Number of them that are misclassified	0	12	4	37
A 103				
Number of predictions that are classified	3	2	3	10
Number of them that are misclassified	0	1	0	7