Quality Specifications for Glucose Meters: Assessment by Simulation Modeling of Errors in Insulin Dose

James C. Boyd* and David E. Bruns

Background: Proposed quality specifications for glucose meters allow results to be in error by 5–10% or more of the “true” concentration. Because meters are used as aids in the adjustment of insulin doses, we aimed to characterize the quantitative effect of meter error on the ability to identify the insulin dose appropriate for the true glucose concentration.

Methods: Using Monte Carlo simulation, we generated random “true” glucose values within defined intervals. These values were converted to “measured” glucose values using mathematical models of glucose meters having defined imprecision (CV) and bias. For each combination of bias and imprecision, 10 000–20 000 true and measured glucose concentrations were matched with the corresponding insulin doses specified by selected insulin-dosing regimens. Discrepancies in prescribed doses were counted and their frequencies plotted in relation to bias and imprecision.

Results: For meters with a total analytical error of 5%, dosage errors occurred in ~8–23% of insulin doses. At 10% total error, 16–45% of doses were in error. Large errors of insulin dose (two-step or greater) occurred >5% of the time when the CV and/or bias exceeded 10–15%. Total dosage error rates were affected only slightly by choices of sliding scale among insulin dosage rules or by the range of blood glucose. To provide the intended insulin dosage 95% of the time required that both the bias and the CV of the glucose meter be <1% or <2%, depending on mean glucose concentrations and the rules for insulin dosing.

Conclusions: Glucose meters that meet current quality specifications allow a large fraction of administered insulin doses to differ from the intended doses. The effects of such dosage errors on blood glucose and on patient outcomes require study.

© 2001 American Association for Clinical Chemistry

Glucose meters and glucose sensors play a central role in the modern management of diabetes. The Food and Drug Administration has approved >25 glucose meters, but quality specifications (or analytical goals) for glucose meters are controversial (1–3). Performance goals for the meters are critically important for clinicians, manufacturers, and most importantly, for patients who wish to achieve optimal control of blood glucose for improved clinical outcomes.

Fraser and Petersen (4) have proposed a hierarchy of criteria for quality specification for analytical methods. For most methods, key determinants of quality specifications are the within-person and person-to-person biological variation of the analyte (4). As Fraser and Petersen state (4), however: “Ideally, quality specifications should be derived objectively from an analysis of medical needs”. An important medical use of glucose meters is in adjustments of insulin dose, with higher doses given at higher glucose concentrations according to predetermined rules for each patient and setting. We felt that it could be useful to explore the possibility of relating quality specifications for glucose meters to this clinical use.

In this study, we asked the question: What is the effect of analytical performance on the ability to correctly direct the administration of the dose of insulin intended for the patient’s (“true”) glucose concentration? We reasoned that a method with high imprecision or bias would frequently yield results sufficiently different from the patient’s true glucose that the insulin dose would differ from the intended dose. To assess this effect, we used simulation modeling to simulate glucose meters with specified bias and imprecision. The modeling provided confident estimates of the dosage error rates for meters of known imprecision and/or bias.
Materials and Methods
We used the SAS package (SAS Institute) to carry out Monte Carlo trials to simulate the effects of specified assay imprecision and bias on insulin dose. We used a clinical sliding scale for administration of insulin for blood glucose concentrations between 150 and 450 mg/dL (8.3 and 25 mmol/L) that was divided into increments of 30 mg/dL (1.67 mmol/L) or 50 mg/dL (2.77 mmol/L). Such scales are frequently used to guide the dosage of insulin in type 1 diabetes. We labeled the intervals as “categories”, and analyzed the prevalence of glucose measurements that corresponded to insulin doses different from the insulin dose that the true glucose concentration called for.

In implementing the computer simulations, we first generated true blood glucose results between 150 and 450 mg/dL (8.3 and 25 mmol/L), using a random number generator following the uniform distribution. Thus, any glucose result in the range had an equal likelihood of occurring. These initial (input) glucose values were considered the true glucose concentration (GlucT). To simulate the effects of analytical imprecision and bias, the true glucose results were modified using the following formula:

$$\text{Gluc}_M = \text{Gluc}_T + [n(0,1) \times \text{CV} \times \text{Gluc}_T] + [\text{Bias} \times \text{Gluc}_T]$$

where

- \( \text{Gluc}_T \) = the true glucose concentration;
- \( \text{Gluc}_M \) = the measured glucose concentration reflecting the effects of analytical imprecision and bias;
- \( \text{CV} \) = the CV of the assay expressed as a fraction;
- \( n(0,1) \) = a random number drawn from a gaussian distribution with a mean of 0 and a SD of 1 to reflect assay imprecision;
- \( \text{Bias} \) = the assay bias (expressed as a fraction).

\( \text{Gluc}_M \) and \( \text{Gluc}_T \) were then assigned to one of the 50 mg/dL (2.8 mmol/L) categories between 150 and 450 mg/dL (8.3 and 25 mmol/L), i.e., \(<200, 200–249, 250–299, 300–349, 350–399, \geq 400 \text{ mg/dL} \approx 11.1, 11.1–13.8, 13.9–16.6, 16.7–19.4, 19.4–22.2, \geq 22.2 \text{ mmol/L}\). The consecutive categories were numbered 1–6, respectively. For convenience, we assumed insulin doses of 0, 2, 4, 6, 8, and 10 units for categories 1–6, respectively. The difference in the category numbers for each \( \text{Gluc}_M \) and \( \text{Gluc}_T \) pair was tallied for each value of assay imprecision and bias. We computed the percentages of observations with (a) zero, (b) one, or (c) two or more category differences (i.e., insulin-dose differences) between the categories (doses).

\(^1\) At the suggestion of a reviewer, glucose values have been revised to be expressed as mg/dL, followed by mmol/L in parentheses. The reverse order is specified for this Journal. Some quality specifications to be discussed here, however, have been expressed only in mg/dL (the converse does not occur), and the program in the Appendix was written assuming mg/dL. Thus, the presentation in the text has been revised to be consistent with the references and the program that are described and discussed.

Sample SAS code to carry out these simulations is listed in the Appendix.

An additional sliding scale for insulin dosing described by Schiffrin and Belmonte (5) was also simulated. In this case, the true glucose values were generated in an interval from 30 to 280 mg/dL (1.7–15.6 mmol/L), and the categories used for insulin administration were <60, 60–90, 90–120, 120–150, 150–200, 200–250, and ≥250 mg/dL (<3.3, 3.3–5.0, 5.0–6.7, 6.7–8.3, 8.3–11.1, 11.1–13.9, and ≥13.9 mmol/L). It should be noted that Schiffrin and Belmonte recommended subtracting two insulin units from the insulin dosage when the glucose concentration was found to be <60 mg/dL, and not changing the insulin dosage for glucose concentrations in the 60–90 mg/dL range. Additional insulin was given only when the blood glucose concentration exceeded 90 mg/dL. In separate simulations, we generated true glucose values that followed a gaussian distribution with a mean (SD) of 163 (35) mg/dL [9.1 (1.9) mmol/L] based on the mean and SD found in actual patient data (6, 7).

Results

Effect of Imprecision with Zero Bias
For the first simulations, we varied the imprecision of the simulated glucose measurements keeping bias = 0. Fig. 1 shows the percentage of cases in which the insulin dose that was appropriate for the simulated glucose result differed from the dose appropriate for the “patient’s” true glucose. The filled columns in Fig. 1 indicate differences of one category (2 units of insulin) or more than one category (4 units of insulin) in insulin dose. The percentages of erroneous doses were greater when the interval size of the categories was decreased from 50 mg/dL (2.8 mmol/L; Fig. 1A) to 30 mg/dL (1.67 mmol/L; Fig. 1B) to test the effect of smaller interval sizes (both interval sizes were used in the Schiffrin/Belmonte scale).

Errors increased with increasing CV of the glucose assay (Fig. 1). As the CV of the glucose assay increased, the percentage of cases in which the measured glucose category differed by two or more categories from the true glucose category became an increasingly higher fraction of the total number of cases. The error rate also was higher when the mean glucose was increased (not shown).

Insulin dose error rates were <5% (i.e., 95% of insulin doses were as intended) only when the CV was ≤1%. Errors of two or more categories (≥4 units of insulin for the model) were exceedingly rare at CVs <5%, but were more common when the CV exceeded 10%. Because two-category (or worse) errors would seem especially undesirable, we paid particular attention to their rates in the following simulations, examining the conditions that kept such errors below 0.2% of the 20 000 modeled insulin doses.
Effect of bias with zero imprecision

With the CV of the glucose assay held constant at 0%, the bias of the assay was varied from 0% to 20%. The percentages of cases in which the measured and true glucose categories differed by one or more than one are plotted in Fig. 2. As the assay bias increased, so did the percentages of cases showing differences in the category assignments of true vs measured glucose. Maintaining error rates for insulin dose, 5% required that the bias be <1%. A bias of 5% maintained two-category errors <0.2%, but the total error rate was 67.5%.

Effect of combined bias and imprecision

To better characterize the combined influence of assay bias and imprecision on the percentage of cases with differences of one category or more than one category in the true vs measured glucose categories, further simulations were carried out to generate contour plots. Fig. 3 shows results with the same sliding scale and glucose concentrations [150–450 mg/dL (8.3–25 mmol/L), uniform distribution] as in Figs. 1 and 2. Fig. 3A shows the total insulin-dosage error rates. Insulin-dosage error rates were <5% when both the CV and bias were <1–1.5%. When both the bias and the CV were 5%, 27% of insulin doses were in error.

The frequency of errors of two or more categories of insulin dose is shown in Fig. 3B. An assay imprecision (CV) of ≥6.5% and bias <5% were required to minimize the overall frequency of errors and to keep the percentage of cases in which the disagreement was two or more categories below 0.2% of the total.

We wished to explore the effect on the model of assuming better control of glucose. We used the Schiffrin sliding scale (5) and assumed glucose concentrations between 30 and 280 mg/dL (1.7 and 15.5 mmol/L) with a uniform (“flat”) distribution.

As shown in Fig. 4A, insulin-dosage error rates were <5% only when both the CV and the bias were <1.5%. When both the CV and the bias were 5%, 21% of insulin doses were in error. As shown in Fig. 4B, the percentage of cases with dose errors of two or more categories (≥4 units of insulin) was <0.2% as long as the CV was <10% and the assay bias was <7% (Fig. 4B), but ~34% of insulin doses were in error (Fig. 4A).

To test the effect of using an underlying gaussian distribution of glucose results rather than a uniform distribution, we performed a simulation using the Schiffrin sliding scale (as in Fig. 4) with glucose concentrations following a gaussian distribution with a mean of 163 mg/dL (9.1 mmol/L) (6) and a SD of 35 mg/dL (1.9 mmol/L) (7). To maintain a frequency of two-category disagreements <0.2% when using the normally distributed data, the glucose monitoring instrument needed to...
Our simulation modeling study relates performance characteristics of glucose analyzers to error rates in insulin dosing. The sliding scale and glucose concentrations were the same as in Figs. 1A and 2. In panel A, 10,000 measurements were simulated at each combination of bias and CV. In panel B, 20,000 measurements were simulated at each combination. Dashed lines and dotted lines indicate the OPSpecs lines for assays with 10% and 5% total error, respectively (14).

Discussion

have bias and imprecision values slightly lower (by <1%) than those found using a uniform distribution of glucose (comparison data not shown).
dosage. These results have implications for the quality specifications (analytical goals) for the performance of these meters. Widely divergent quality specifications for analytical performance have been set by various approaches, including expert opinion (8, 9), opinion of clinicians (10), government regulation (11), and biological variation (3). Of these approaches, the goals based on biological variation appear to be the most stringent. Proponents of this approach suggest that assay imprecision should not exceed one-half of the within-subject biological variation (12). Thus, for blood glucose using this approach, a CV <2.2% has been suggested as a target for within-laboratory analytical imprecision (3). Our modeling shows that at this CV, nearly 10% of insulin doses will be in error even if the bias is zero.

Multiple goals for self-monitoring of blood glucose devices have been proposed. In 1987, the American Diabetes Association (ADA) recommended that measured glucose concentrations in the range of 30–400 mg/dL differ by <10% from the true glucose concentration (8). Subsequently, the ADA revised the performance goal to 5% (9). The Clinical Laboratory Improvement Amendments of 1988 (CLIA) specify that meters be within 10% of target values or ± 6 mg/dL, whichever is larger (11). Clarke et al. (10) proposed the method of Error Grid Analysis, which identifies clinically important errors by use of broad target ranges on a graph.

Our study takes a different approach to quality specifications for glucose meters by quantifying the effects of meter bias and imprecision on their use with a sliding insulin scale to identify the insulin dose appropriate for the (true) blood glucose. The results indicate that glucose assay imprecision (CV) <6–10% combined with bias <5–7% will only infrequently (<0.2%) lead to insulin dosage errors of more than one category, but will allow ~25–34% of insulin doses to differ from the intended doses (Figs. 3 and 4).

The results of the simulation modeling can be used to assess the effect on insulin dosing of the bias and imprecision reported for glucose meters. Weitgasser et al. (13) recently compared four newer glucose meters with four older ones. Because an experienced technician performed all measurements, the results may be taken as an indication of the optimum performance achievable by the meters. For three of the four newer meters, the CVs were consistently <5% at each of three concentrations tested. One meter achieved CVs <3.5%; another had all CVs <2%. The older meters achieved similar CVs at midrange and high concentrations of glucose, but were imprecise (CVs of 7–16%) at low glucose concentrations (2.9–3.9 mmol/L) below the concentrations of interest for insulin-dosage adjustments. The bias of newer meters was markedly improved, with a mean absolute bias (expressed as a percentage) for the newer meters of 1.7% (range, 1.1–2.2%) vs 10% (range, 5.8–13.5%) for the older meters. For the older meters, our results show that ~40% of insulin doses are in error as a result of the 10% bias. Using the mean values of the current generation of devices for imprecision (2.8% in the range of interest) and bias (1.7%), our simulation modeling predicts a total error rate of ~10% and errors of more than one dose category of <0.2%. The performance of meters in the hands of patients, however, is unlikely to be as good as the values achieved in the study by Weitgasser et al. (13).

Equations for relating bias and imprecision of assays to total error goals, such as those designated by CLIA, the ADA, and NCCLS, have been published by Westgard (14). The OPSpecs chart defined by these equations (when no quality control is being used) for a 10% criterion (CLIA88 and ADA87) may be used to define what bias and precision combinations for an assay will satisfy that criterion. With this approach, assuming a z value of 1.68 (95% confidence that results will meet the criterion) and superimposing the OPSpecs line on our Figs. 3A and 4A, we find that for a 10% total-quality goal, between 16% and 45% of insulin doses will be in error. With a 5% total-quality goal (ADA96), between 8% and 23% of insulin doses will be in error. The error rates for two-category errors with either the 5% or 10% total error criterion are <0.2%.

The simulation modeling described here has several advantages. It is simple: when used with the Monte Carlo features of a program such as SAS, only a few lines of programming are needed (see the Appendix). The trials are fast, requiring <30 min on a modest desktop computer to simulate 10,000 measurements at each of 400 combinations of bias and imprecision. The approach can be readily modified to test other insulin dosage schedules (“sliding scales”), to model different assumptions about the distribution of glucose measurements, or to use actual glucose measurements downloaded from a patient’s meter.

This study has several limitations. It is, of course, a simulation, but the ~25 million measurements described here could not be accomplished in a real trial of less than several years’ duration and involving thousands of individuals. Moreover, the problem is ideally suited to modeling because the consequences (insulin dosage errors) follow logically and inescapably from the starting conditions (meter bias and imprecision, glucose values, and insulin dosage schedules). In addition, we have examined only selected sliding scales. The scales chosen were, however, markedly different, and the results were affected only slightly by choice of scale. Importantly, the modeling did not address patient outcomes. Conceivably, computer modeling can be used to relate insulin-dosage errors to changes in mean blood glucose, which can be related to complications based on data such as those of the Diabetes Complications and Control Trial. Such modeling is, however, beyond the scope of this project. Finally, the study does not address the importance of frequency of glucose measurement and of insulin-dosage adjustment. When frequency is high (e.g., several times per hour), random errors are probably of little consequence. In this
case, however, a plot showing the effects of bias (e.g., Fig. 2) can be useful.

In conclusion, simulation modeling indicates that glucose meters that achieve both a CV and a bias <5–6% rarely lead to major errors in insulin dose, but the CV and bias must be <1–1.5% to keep rates of smaller errors below 5%. Although some existing meters have the capability of producing results of this quality, it is less clear whether existing meters achieve this level of performance in the hands of all patients. Efforts in meter design may need to focus less on improving the precision and trueness that can be obtained under ideal conditions and more on continuing the trend toward meters that produce quality results in the hands of users who may have special needs.

Appendix

The appendix for this report is available as an electronic supplement from the Clinical Chemistry Web site. The file can be accessed by a link from the online Table of Contents (http://www.clinchem.org/content/vol47/issue2).

This project was conducted with no outside support. We thank Dr. Eugene Barrett, University of Virginia, for helpful discussions.

References