The Danger of Using Total Error Models to Compare Glucose Meter Performance

Jan S. Krouwer, PhD[1]

Abstract

Glucose meter performance specifications provide limits for 95% of results, which is the same as total error. A popular total error model is that total error equals (average) bias plus 2 times imprecision. This model has been used to specify combinations of average bias and imprecision that satisfy total error goals. But this model is incomplete and its conclusions are suspect. It is shown that when interferences occur in glucose meters as exemplified by hematocrit interference, the total error model proposed by Boyd and Bruns cannot distinguish between meters that differ in performance. The CLSI standard EP21-A, does not have this problem because it directly estimates total error bypassing the need for a model. An example illustrates these points.

Keywords

glucose meter, interferences, performance standard, random bias, simulation, total error

With the availability of so many glucose meters, one needs to know what acceptable performance is. Two standards organizations have addressed this with performance limits for glucose meters.[1-2] Although I have commented that these standards fail to provide limits for 100% of the results,[3] the limits provided for 95% of the data are an important criterion for glucose meter quality. Westgard stated that total error, which is represented by the location of 95% of the data, is of prime importance to clinicians.[4] Moreover, he developed a simple model to estimate total error as:

$$\text{Total error} = \text{bias} + 2 \times \text{SD} \qquad \text{Equation 1}$$

Thus, total error equals bias plus imprecision. This model is intuitively appealing, since what else could there be besides bias and imprecision. Boyd and Bruns have used this model to show combinations of bias and imprecision needed to keep total error within limits for glucose meters.[5]

The purpose of this article is to show that this model is incomplete and how it can mislead one in estimating glucose meter performance. First, it is noted that in the Westgard model, what is meant by bias is really *average* bias of a series of specimens. Lawton and coworkers provided a more complete model to estimate total error.[6] Their model adds a random bias term (as a standard deviation) to the Westgard model. This additional term accounts for interferences that vary from sample to sample. A problem with the Lawton model is that the extra term is difficult to estimate.

This random bias term is sometimes thought to deal with large, rare interferences, but it accounts for any size of interference, and this is especially pertinent to glucose meters. Thus, some glucose meters suffer from hematocrit interference and

others not.[7] The CLSI standard EP21-A takes a different approach to estimating total error by directly computing the differences between the candidate assay and reference.[8] Thus, no modeling is required. The difference between EP21-A and the Boyd and Bruns method can be shown by comparing 2 hypothetical glucose meters, A and B. As shown in Table 1, the 2 meters both have no statistically significant average bias and the same precision. But meter B shows 20% bias at the extremes of hematocrit.

According to the Boyd and Bruns model, glucose meters A and B have the same total error because they have the same average bias and precision (equation 1). But when analyzed with a CLSI EP21-A mountain plot,[8-9] meter B with hematocrit interference is clearly not as accurate as meter A and fails the POCT12-A3 glucose meter standard (Figures 1-2). In a mountain plot, the glucose differences from reference are sorted from low to high and ranked. The Y axis represents the cumulative probability which normally ranges from 0 to 1. But to present a plot that is easier to visualize, the mountain plot cumulative probability values above 0.5 have been subtracted from 1 to give adjusted values. Two worked examples of how to construct a mountain plot using a spreadsheet are explained in EP21-A.

This demonstration was performed by simulations and simulations always work. The hematocrits were chosen as discrete values uniformly spanning 32% to 56% and applied
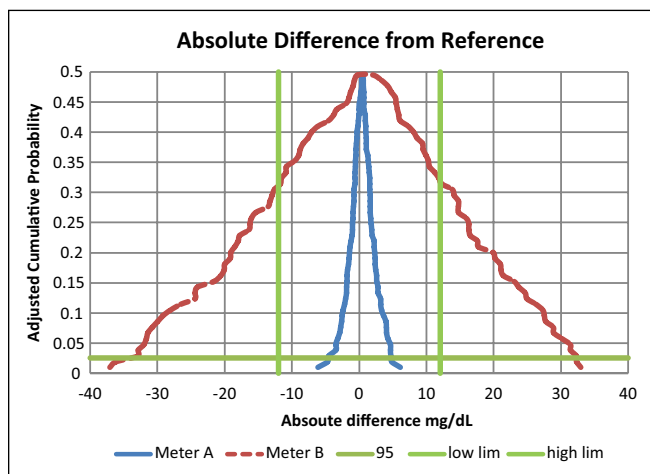
[1]Krouwer Consulting, Sherborn, MA, USA

Corresponding Author:
Jan S. Krouwer, PhD, Krouwer Consulting, 26 Parks Dr, Sherborn, MA 01770, USA.
Email: jan.krouwer@comcast.net

Table 1. Performance Attributes of Two Glucose Meters.

Meter	Average bias	Precision (CV)	Hematocrit interference
A	0	5%	None
B	0	5%	+20% low hematocrit, −20% high hematocrit

Figure 1. Absolute differences for glucose meters for reference less than 100 mg/dL. The intersections of the horizontal and the straight vertical lines represent the limits to contain 95% of the data. A meter that is contained with this space meets goals. Thus meter A meets goals and meter B does not.
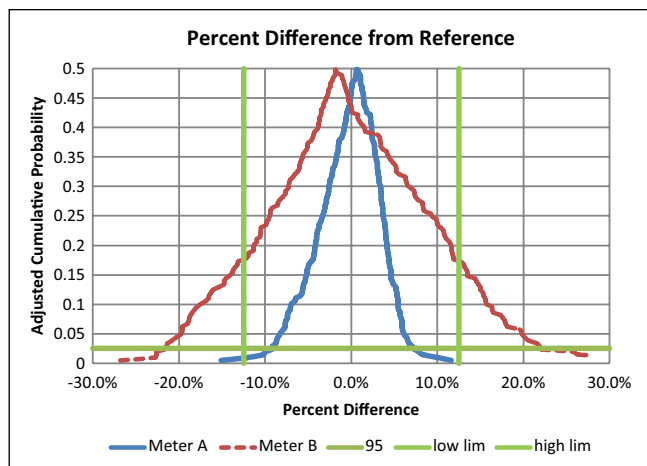
Figure 2. Percentage differences for glucose meters for reference greater than 100 mg/dL. The intersections of the horizontal and the straight vertical lines represent the limits to contain 95% of the data. A meter that is contained with this space meets goals. Thus meter A meets goals and meter B does not.

to discrete glucose values uniformly spanning 30 to 280 mg/dL. Had different simulation conditions been used, meter A would remain the same and meter B might have become more narrow or wider in Figures 1 and 2. Only hematocrit

interference was chosen. Other interfering substances would widen a meter's total error performance and at the same time not be detected by the Boyd and Bruns approach. The fact that the average bias is not statistically significant is because manufacturers calibrate their systems to guarantee this property.

Yet, this commentary is not the first objection to the Boyd and Bruns model. I critiqued their model in a fashion similar to this commentary and they responded.[10-11] In their response, they said I was correct but the sources of error I mentioned were "outside the scope of our study, in part because it is difficult to know how one might model the interferences." They went on to say that in their article they discussed the need for manufacturers to "design instruments that avoid sources of error, such as those encountered by patients with special needs." Unfortunately, my critique had no effect because their model continues in recent articles as if the critique never happened.[12-13] Moreover, in the recently released CLSI glucose meter standard, POCT12-A3,2 these models are cited as a basis for the performance limits for glucose meters. Ironically, Boyd and Bruns[11] state in their response to my critique: "The points raised in Dr. Krouwer's letter do point out that our estimates of quality requirements, as demanding as they may seem, would become even more demanding if the additional sources of error were included." In a similar story, I critiqued[14] the NCEP's use of the Westgard model to arrive at performance goals for cholesterol.[15] In spite of objections, the Westgard model also persists.[16] Perhaps these models persist *because* they are models and (simple) models are satisfying. In a total error analysis conducted using CLSI EP21-A, there is no means to separate error components nor a basis for setting limits on error components.

Finally, it is noted that total error only captures error that is allowed to occur in the experiment. For example, such experiments are often done with a single lot of reagent with many conditions controlled more tightly that would occur in routine use.

Abbreviations

CLSI, Clinical Laboratory Standards Institute; NCEP, National Cholesterol Education Program.

Declaration of Conflicting Interests

Funding

References

1. *ISO 15197:2013. In Vitro Diagnostic Test Systems— Requirements for Blood-glucose Monitoring Systems for*

Self-testing in Managing Diabetes Mellitus. Geneva, Switzerland: International Organization for Standardization; 2013.

2. *CLSI POCT12-A3 Point-of-care Blood Glucose Testing in Acute and Chronic Care Facilities; Approved Guideline*. 3rd ed. Wayne, PA: CLSI; 2013.

3. Krouwer JS. Why specifications for allowable glucose meter errors should include 100% of the data. *Clin Chem Lab Med*. 2013;51:1543-1544.

4. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem*. 1974;20:825-833.

5. Boyd JC, Bruns DE. Quality specifications for glucose meters: assessment by simulation modeling of errors in insulin dose. *Clin Chem*. 2001;47:209-214.

6. Lawton WH, Sylvester EA, Young-Ferraro BJ. Statistical comparison of multiple analytic procedures: application to clinical chemistry. *Technometrics*. 1979;21:397-409.

7. Brazg RL, Klaff LJ, Parkin CG. Performance variability of seven commonly used self-monitoring of blood glucose systems: clinical considerations for patients and providers. *J Diabetes Sci Technol*. 2013;7:144-152.

8. *CLSI EP21-A Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline*. Wayne, PA: CLSI; 2013.

9. Krouwer JS, Monti KL. A simple graphical method to evaluate laboratory assays. *Eur J Clin Chem Clin Biochem*. 1995;33:525-527.

10. Krouwer JS. How to improve total error modeling by accounting for error sources beyond imprecision and bias. *Clin Chem*. 2001;47:1329-1330.

11. Boyd JC, Bruns DE. Drs. Boyd and Bruns respond. *Clin Chem*. 2001;47:1330-1331.

12. Karon BS, Boyd JC, Klee GG. Glucose meter performance criteria for tight glycemic control estimated by simulation modeling. *Clin Chem*. 2010;56:1091-1097.

13. Boyd JC, Bruns DE. Monte Carlo simulation in establishing analytical quality requirements for clinical laboratories tests meeting clinical needs. *Methods Enzmol*. 2009;467:311-433.

14. Problems with the NCEP (National Cholesterol Education Program) Recommendations for Cholesterol Analytical Performance. *Arch Pathol Lab Med*. 2003;127:1249.

15. NCEP Recommendations on Lipoprotein Measurement. NIH publication no. 95-3044. Available at: http://www.nhlbi.nih.gov/health/prof/heart/chol/lipoprot.pdf. Accessed October 23, 2013.

16. Westgard JO, Westgard SA. Total analytic error. From concept to application. *Clin Chem News*. 2013;39:8-10. Available at: http://www.aacc.org/publications/cln/2013/ september/Pages/Total-Analytic-Error.aspx#. Accessed October 23, 2013.