The Need for Clinical Accuracy Guidelines for Blood Glucose Monitors

David C. Klonoff, M.D., FACP

The performance of blood glucose (BG) monitors can be classified based on analytical accuracy or clinical accuracy.[1] Analytical accuracy represents a quantitative approach to describing how closely the result of a measurement method being evaluated compares with a measurement by a reference method. Clinical accuracy is a qualitative approach that describes the clinical outcome of basing a treatment decision on the result of a measurement method being evaluated.[2] Analytical accuracy is measured by various statistical metrics, including precision and bias, among others. Clinical accuracy is currently measured by comparing paired data points of the results of the measurement being evaluated along with the results from a reference method, with the paired data points plotted on a grid, known as an error grid.

On an error grid, each data point (representing both the BG monitor value and the reference value in two dimensions) can be mapped out as lying within a performance zone. This approach permits data sets to be defined on the basis of the percentage of data points that fall into each zone or category of clinical outcome. The error grid method assigns each data point a performance zone based on whether there is an effect on the clinical action and if so, then how it will affect the clinical outcome. Most analytical accuracy guidelines for BG monitors account for less than 100% of data points comparing a test method with a reference method, whereas an error grid always accounts for 100% of the data points.

International Organization for Standardization (ISO) analytical accuracy standard 15197 for BG monitors specifies that 95% of data points must demonstrate acceptable analytical accuracy but does not specify any performance targets for the remaining 5% of data points. A generic error grid, which would pertain to clinical accuracy of BG monitors, typically contains at least three zones (**Figure 1**).[3] This approach to creating an error grid containing at least three zones is discussed in Clinical Laboratory Standards Institute EP-27, which is in development.[4] Each of these three basic zones could then be divided into multiple zones of relatively better and relatively worse performance, so that an error grid of clinical accuracy can have more than three zones, but ultimately, the zones describe performance that is either very acceptable, barely acceptable, or unacceptable.
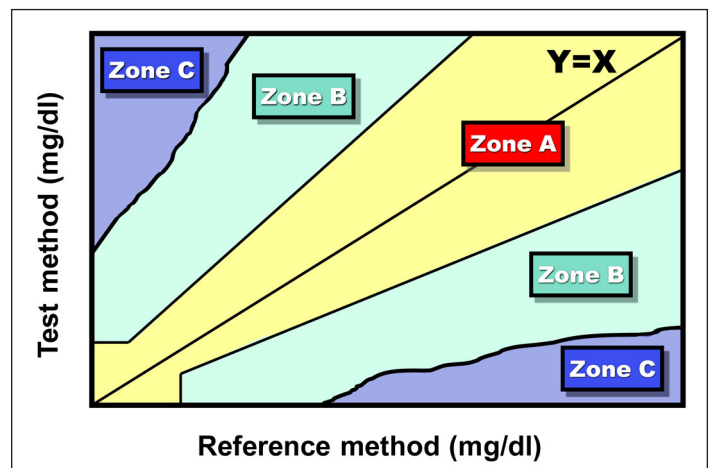

Figure 1. A generic error grid comparing a test method with a reference method. Zone A is the allowable total error region where at least 95% of the data points should be found. Zone C is the limits of erroneous region where no data should be found. Zone B is in between Zones A and C where no more than 5% of the data points should be found (Adapted from Reference 3).

Author Affiliation: Mills-Peninsula Health Services, San Mateo, California

Abbreviations: (BG) blood glucose, (CGM) continuous glucose monitor, (DCCT) Diabetes Control and Complications Trial, (FDA) Food and Drug Administration, (ISO) International Organization for Standardization

Keywords: accuracy, blood glucose monitor, error grid, FDA, ISO, statistic

Corresponding Author: David C. Klonoff, M.D., FACP, Mills-Peninsula Health Services, 100 South San Mateo Dr., Room 5147, San Mateo, CA 94401; email address *dklonoff@diabetestechnology.org*

Two types of error grids are used for describing the clinical accuracy of BG monitors. One is the Clarke error grid, which was developed by five diabetes experts from the University of Virginia and published in 1987.[5] The other type is the Parkes error grids—two grids also known as the consensus error grids—which were intended for type 1 diabetes and type 2 diabetes.[6] They were developed by four diabetes experts from Becton Dickinson, Inc., and Albert Einstein College of Medicine. These two error grids were developed in 1994 through a survey of 100 physicians who treat patients with diabetes and who were attending the 1994 American Diabetes Association Meeting. These error grids were published 6 years later in 2000. The Parkes error grids retained the five-risk zone format of the Clarke error grid, but this consensus metric utilized an expert panel (instead of a team of five authors) to draw exact boundaries between zones. The Parkes error grid was crafted to eliminate several discontinuities in the Clarke error grid, where an infinitesimal change in the BG level caused the risk category to increase by two or even three levels of risk.

An error grid has been proposed for use as a tool in describing the point accuracy of continuous glucose monitors (CGMs).[7] This statistic does not account for rate of change information, and it might produce an inflated notion of the true accuracy of a CGM.[8] A modification of the error grid used for evaluating BG monitor clinical accuracy, known as continuous glucose-error grid analysis, has been proposed to serve as a method of evaluating the clinical accuracy of CGMs by accounting for both the accuracy of the glucose values as well as the direction and rate of change of the BG fluctuations.[9]

Other error grids for quantifying the clinical performance of BG monitors have been proposed. A team of 10 hospital experts from Austria and the Czech Republic has proposed an insulin titration error grid with paired data points assembled into zones according to how severely the error would impact a decision for how much insulin to administer to a hospital patient. A measurement error of ±20% was considered a tolerable level of error, which was the most clinically accurate zone in this system.[10] A team of two veterinarians from Scotland and the United Kingdom has proposed an error grid to analyze clinical accuracy of BG monitors that are used for cats. A measurement error of ±10% was considered a tolerable level of error.[11] Error grids for investigational point-of-care prothrombin measurement devices have been proposed to compare the performance of test methods with reference methods and generate zones of clinical accuracy.[12–14]

Blood glucose monitors are approved by the Food and Drug Administration (FDA) if they meet analytical accuracy criteria as defined by ISO 15197, which is entitled "*In vitro* diagnostic test systems. Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus." Currently, this standard is being revised. The FDA has also announced plans to revise their own guidelines for accuracy of BG monitors, which will include new requirements for analytical accuracy.[15]

The benefit of including a clinical accuracy requirement in the guidelines would be that such a metric can be used to describe outlier data points from BG monitors that are not acceptable analytically but that might or might not be acceptable clinically. These data points can be classified according to whether their degree of inaccuracy will lead to untoward clinical consequences and if so, then how severe the consequences might be.

Error grids can, therefore, be used to classify the seriousness of outlier data points that result in altered clinical action. This classification can be useful for determining whether to grant regulatory clearance of specific BG monitors that perform acceptably based on their clinical accuracy. At this time, there is no generally agreed upon standard (irrespective of which error guide is used) for any error grid as to what percentage of data points must fall into the highest performance zone or which zones are to be defined as providing adequate clinical accuracy. This type of tool can potentially ensure safe performance of BG monitors if a regulatory body allows only highly performing monitors, with respect to clinical accuracy, to be on the market.[16]

If the FDA elects to begin using a formal metric of clinical accuracy as a factor in determining regulatory approval of BG monitors, then it will be necessary for a new clinical accuracy metric to be developed because the two most widely used error grids have become obsolete for three reasons.

First, the more recent of the two types of error grids will soon be 18 years old, and it was developed less than 1 year after publication of the Diabetes Control and Complications Trial (DCCT) study, the first clinical trial to demonstrate the benefits of intensive glycemic control.[17] The older error grid was developed prior to publication of the DCCT. Second, both types of error grids were developed prior to the introduction of analog insulins, which have allowed more intensive insulin therapy in response to information from BG monitors.[18] Third, the current error grids were developed in an era

where acceptable analytical accuracy for the majority of data points was defined as a difference between the test method and the reference method of ±20%. This difference was used to define the border between ideal clinical accuracy and suboptimal clinical accuracy in the Clarke error grid and appears to have influenced how the border was specified in the Parkes error grids. The next ISO guideline and the next FDA guideline are both expected to specify a tighter range of accuracy than ±20% for acceptable analytical accuracy, which will render the existing error grid borders between ideal clinical accuracy and suboptimal clinical accuracy as obsolete.

These two existing error grid types do not distinguish between clinical states where tight glycemic control is sought and where glycemic goals are less strict. Although the Parkes error grid was developed to provide separate metrics for type 1 and type 2, the type 2 error grid has fallen out of use. At this time, no widely applied metric for clinical accuracy of a BG monitor mandates a narrower range of adequate clinical accuracy for states requiring tight control. Some clinical conditions might impact both the target ranges for clinical accuracy and the magnitude of tolerable analytical inaccuracy with either type of diabetes. These conditions could include hospitalization in an intensive care unit, hospitalization in a ward, outpatient insulin pump therapy, type 1 diabetes on multiple dose insulin therapy, hypoglycemia unawareness, pregnancy, and corticosteroid therapy.

If a new error grid is to be developed, then a variety of questions must be addressed. If consensus can be reached on the most important elements of a metric for clinical accuracy of BG monitors, then the new error grid will be well grounded in current clinical practices as well as the needs of both clinicians and regulators. At least 15 issues related to error grids must be considered in the process of developing a new error grid. These issues are listed in **Table 1**.

Two standards dealing with BG monitor performance are expected to be published in 2012. These standards are listed in **Table 2**.[19] One focuses on monitoring in the outpatient setting and the other focuses on monitoring in hospitals and long-term facilities. These two documents will likely provide guidance to the FDA on setting new standards for BG monitor performance.

It is becoming clear that defining only analytical accuracy of BG monitors is not sufficient. Clinical accuracy must also be considered to fully assess their performance. Regulatory agencies need a metric for this purpose to

Table 1.
Issues Related to Developing a New Error Grid for Clinical Accuracy of Blood Glucose Monitor Performance

1	Is a metric of clinical accuracy needed for BG monitors?
2	What is the clinical significance of outlier data points?
3	What are the advantages and disadvantages of both currently used error grids?
4	How will impending tighter analytical accuracy standards affect the usability of currently used error grids?
5	Is a new error grid the best tool for defining clinical accuracy of BG monitors?
6	How many different error grids are needed for various clinical states, including hospital settings?
7	How many zones are appropriate for defining different levels of clinical performance?
8	Should the border between acceptable clinical accuracy and lesser levels of clinical accuracy be based on an absolute difference or a relative difference between a test method and a reference method?
9	Should the ideal zone be a single zone of acceptable paired data points or else multiple smaller zones of increasingly better-than-required levels of clinical performance?
10	Should the same criteria for clinical accuracy acceptance be applied across the range of reference BG values?
11	Should a consensus of experts be used in the process of setting clinical accuracy standards?
12	What should be the criteria for defining, identifying, and contacting experts for participation in a consensus metric?
13	How can the opinions of multiple experts be combined into a structured tool?
14	How many data points and what type of glycemic distribution in a study are needed to assess the clinical accuracy of a BG monitor using an error grid?
15	What distribution of clinical accuracy performance outcomes as defined by an appropriate error grid should be required for regulatory approval of a BG monitor?

Table 2.
Standards for Blood Glucose Monitor Performance That Are Expected to be Released in 2012

1. ISO 15197: *In vitro* diagnostic test systems—requirements for blood glucose monitoring systems for self-testing in managing diabetes mellitus. From the International Organization for Standardization.

2. POCT12-A3: *Point-of-Care Blood Glucose Testing in Acute and Chronic Care Facilities; Approved Guideline*— Third Edition (Formerly C30-A2). From the Clinical and Laboratory Standards Institute.

apply to data sets if they are going to define the clinical accuracy of BG monitors. Now is the time to define the clinical accuracy of BG monitors by developing a new error

grid as a modern metric that reflects current clinical practices, currently available types of insulin, and current analytical accuracy standards.

References:

1. Krouwer JS, Cembrowski GS. A review of standards and statistics used to describe blood glucose monitor performance. J Diabetes Sci Technol. 2010;4(1):75–83.

2. Boren SA, Clarke WL. Analytical and clinical performance of blood glucose monitors. J Diabetes Sci Technol. 2010;4(1):84–97.

3. Clinical Laboratory Standards Institute. EP27-P. Vol. 29, No. 16. How to construct and interpret an error grid for diagnostic assays; proposed guideline. http://www.clsi.org/source/orders/free/ep27-p.pdf. Accessed December 6, 2011.

4. http://www.clsi.org/Content/NavigationMenu/Committees/EvaluationProtocols/ProjectsinDevelopment/Projects_in_Developm.htm. Accessed December 6, 2011.

5. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care. 1987;10(5):622–8.

6. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. Diabetes Care. 2000;23(8):1143–8.

7. Garg SK, Potts RO, Ackerman NR, Fermi SJ, Tamada JA, Chase HP. Correlation of fingerstick blood glucose measurements with GlucoWatch biographer glucose results in young subjects with type 1 diabetes. Diabetes Care. 1999;22(10):1708–14.

8. Kollman C, Wilson DM, Wysocki T, Tamborlane WV, Beck RW. Diabetes Research in Children Network Study Group. Limitations of statistical measures of error in assessing the accuracy of continuous glucose sensors. Diabetes Technol Ther. 2005;7(5):665–72.

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 data. Diabetes Care. 2004;27:1922–8.

10. Ellmerer M, Haluzik M, Blaha J, Kremen J, Svacina S, Toller W, Mader J, Schaupp L, Plank J, Pieber T. Clinical evaluation of alternative-site glucose measurements in patients after major cardiac surgery. Diabetes Care. 2006;29(6):1275–81.

11. Dobromylskyj MJ, Sparkes AH. Assessing portable blood glucose meters for clinical use in cats in the United Kingdom. Vet Rec. 2010;167(12):438–42.

12. Hemkens LG, Hilden KM, Hartschen S, Kaiser T, Didjurgeit U, Hansen R, Bender R, Sawicki PT. A randomized trial comparing INR monitoring devices in patients with anticoagulation self-management: evaluation of a novel error-grid approach. J Thromb Thrombolysis. 2008;26(1):22–30.

13. Torreiro EG, Fernández EG, Rodríguez RM, López CV, Núñez JB. Comparative study of accuracy and clinical agreement of the CoaguChek XS portable device versus standard laboratory practice in unexperienced patients. Thromb Haemost. 2009;101(5):969–74.

14. Petersen JR, Vonmarensdorf HM, Weiss HL, Elghetany MT. Use of error grid analysis to evaluate acceptability of a point of care prothrombin time meter. Clin Chim Acta. 2010;411(3–4):131–4.

15. Klonoff DC. The Food and Drug Administration is now preparing to establish tighter performance requirements for blood glucose monitors. J Diabetes Sci Technol. 2010;4(3):499–504.

16. Klonoff DC. Improving the safety of blood glucose monitoring. J Diabetes Sci Technol. 2011;5(6):1307–11.

17. The Diabetes Control and Complications Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med. 1993;329:977–86.

18. [No authors listed]. Rapid-acting insulin approved for marketing. Am J Health Syst Pharm. 1996;53(19):2250.

19. American Association for Clinical Chemistry. Blood glucose meters: is FDA ready to tighten up accuracy standards? Clinical Laboratory News. 2010;36(5). http://www.aacc.org/publications/cln/2010/may/Pages/CoverStory1May2010.asp. Accessed December 8, 2011.