Journal of Diabetes Science and Technology
Volume 7, Issue 6, November 2013

REVIEW ARTICLE

Assessing the Analytical Performance of Systems for Self-Monitoring of Blood Glucose: Concepts of Performance Evaluation and Definition of Metrological Key Terms

Oliver Schnell, M.D.,[1] Rolf Hinzmann, M.D., Ph.D.,[2] Bernd Kulzer, Ph.D.,[3] Guido Freckmann, M.D.,[4] Michael Erbach, M.D.,[5] Volker Lodwig, Ph.D.,[2] and Lutz Heinemann, Ph.D.[6]

Abstract

Reliability of blood glucose (BG) measurements is a prerequisite for successful diabetes management. Publications on the evaluation of self-monitored glucose values, however, are frequently characterized by a confusion in terminology. We provide an inventory of key terms such as accuracy, trueness, precision, traceability, calibration, and matrix effect to avoid future misunderstanding. Definitions are taken from the metrological literature and international norms and explained in a language intended for nonspecialists in metrology. The terms are presented in light of the need to apply generally accepted definitions. In addition, a description of requirements and components for a sound evaluation of BG measurement systems is presented. These factors will also enable improvement in future comparisons of study results.

J Diabetes Sci Technol 2013;7(6):1585–1594

Introduction

The use of blood glucose (BG) systems for self-monitoring of blood glucose (SMBG) both in insulin-treated and non-insulin-treated people with diabetes is supported by trials, reviews, meta-analyses, and guidelines.[1–10] A BG system is the combination of a BG meter and test strips. Both of these determine the analytical performance of a given BG system. Self-monitoring of blood glucose is reported to support preventive strategies against acute and chronic complications of diabetes, to increase patients' awareness of hypoglycemic symptoms, and therefore to trigger patient-initiated prevention of significant hypoglycemic episodes.[11–13]

This points to the need for high and reliable measurement quality. What are the key terms to describe such a quality? In evaluating glucose measurement quality, which requirements need to be satisfied?

One annoying and rather problematic aspect with regard to all discussions about BG measurement and its related topics is the confusion with terms. In most publications, terms are used with variable definitions. For example, the term

Author Affiliations: [1]Forschergruppe Diabetes e.V., Helmholtz Center Munich, Munich-Neuherberg, Germany; [2]Roche Diagnostics GmbH, Mannheim, Germany; [3]Research Institute of the Diabetes Academy Mergentheim, Bad Mergentheim, Germany; [4]Institute for Diabetes-Technology GmbH, University of Ulm, Ulm, Germany; [5]Sciarc GmbH, Baierbrunn, Germany; and [6]Science & Co GmbH, Düsseldorf, Germany

Abbreviations: (BG) blood glucose, (ID-GC/MS) isotope dilution gas chromatography mass spectrometry, (ISO) International Organization for Standardization, (SMBG) self-monitoring of blood glucose

Keywords: accuracy, blood glucose meter evaluation, diabetes, self-monitoring of blood glucose, trueness

Corresponding Author: Oliver Schnell, M.D., Forschergruppe Diabetes e.V., Helmholtz Center Munich, Ingolstaedter Landstrasse 1, 85764 Munich-Neuherberg, Germany; email address *oliver.schnell@lrz.uni-muenchen.de*

"accuracy" is sometimes used instead of "trueness."[14,15] We therefore propose that all people discussing this topic use a common language that is also used in other areas of laboratory diagnostics to avoid confusion. Subsequently, the aim of this article is to provide an inventory of key terms and to describe in brief the requirements for a sound evaluation of BG systems. This is also of importance in light of the revised International Organization for Standardization (ISO) guidelines. As metrological definitions are sometimes rather difficult to understand for nonspecialists, the definitions are complemented by a description in an easier language where appropriate. In doing so, we had to find a middle ground where the definition might sometimes not give complete justice to a concept, but at the same time, it would also avoid oversimplification.

Terminology

The analytical performance of devices used for *in vitro* diagnostic measurements can be characterized by different performance criteria such as accuracy, bias, and precision. Further, a measurement procedure with which BG systems are compared can be any routine laboratory method or a reference method in a more strict sense. It is important that the method with which BG systems are compared is traceable to a reference standard. As mentioned earlier, statements about this topic, however, are often affected by misunderstandings and inaccuracies in terminology.

Metrology is the science of measurement, and the internationally agreed upon nomenclature of this science is summarized in the *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*.[16] In metrology, the concept of measurement error plays an important role. It is a common although simplified approach to split the so-called total error of measurement into two portions, which are considered independently: random error and systematic error. Random error represents the "noise" of the system and causes deviations in different directions; it is a feature of the technology used. Systematic error, however, as the name implies, is usually constant for all samples and points in one direction. In this case, systematic error can be compensated by calibration while random error cannot. More complex error models consider further error components[17] such as drift, which can be described as a change of bias over the course of time, or random interference. This, however, may go beyond the scope of this review. In the following explanations, we aim to discuss the contributions of systematic error and random error to the total error of measurement.

Both components, random and systematic error, add up (not in a strictly arithmetic way, discussed later) to the total error. The total error is what is clinically important, as it is responsible for potentially wrong clinical decisions.

Random Error, Precision, and Standard Deviation (Coefficient of Variation)

The random error of a method is assessed by determining its precision. According to ISO 5725-1, precision means the closeness of agreement between independent test results obtained under stipulated conditions.[14] Precision is usually described in terms of standard deviation or coefficient of variation, which is the standard deviation as a percentage of the mean.

As this definition of random error is valid irrespective of whether the results are right or wrong, high precision should not be equated with good performance (**Figure 1**). Measurements performed under identical conditions over a short period of time by the same operator on the same sample (or similar samples, e.g., different vials of the same control material) on the same BG system assess the so-called repeatability (sometimes called within-run precision). Measurements performed by different operators at different locations using different BG systems of the same type on similar samples assess the so-called reproducibility (sometimes called laboratory-to-laboratory precision).

Systematic Error, Trueness, and Bias

The systematic error of a method is assessed by determining its trueness. According to ISO 5725, trueness is the closeness of agreement between the average value obtained from a large series of repetitive tests and an accepted reference value.[14] Trueness is usually expressed in terms of bias.[14] The systematic error of a method may be composed of one or more systematic error components.[18] A true mean value can be obtained from a less precise measurement series (**Figure 1**) if the number of measurements is high enough.

Total Error and Accuracy

Total error is a composite of systematic and random error. The total error of a method is assessed by determining its accuracy. Accuracy is defined as the closeness of agreement between a test result and the accepted reference value.[14]

Table 1 summarizes the terminology used for describing analytical error.

The term "accuracy (of the mean)" is sometimes used instead of "trueness."[14,15]

Table 1.
Overview of the Terminology Used for Describing Analytical Error

Type of error	Random error
Performance characteristics	Precision (calculated as standard deviation or coefficient of variation)
Type of error	Systematic error
Performance characteristics	Trueness (calculated as bias)
Type of error	Total error
Performance characteristics	Accuracy

How Do Random Error and Systematic Error Add Up to Total Error?

Contrary to what one might intuitively assume, total error is not simply the arithmetic sum of random error and systematic error, as both components might add up or partly compensate each other. The equation is

$$\text{total error} = z * \text{random error} + \text{systematic error}$$

The z-multiplier is typically chosen at the 95% probability level. At this level, it ranges between 1.65 and 1.96, depending on the ratio between systematic error and random error.[18] If the systematic error is zero, the total error is 1.96 times the random error at this probability level.

High accuracy is possible only in the presence of high precision as well as high trueness (**Figure 1**).

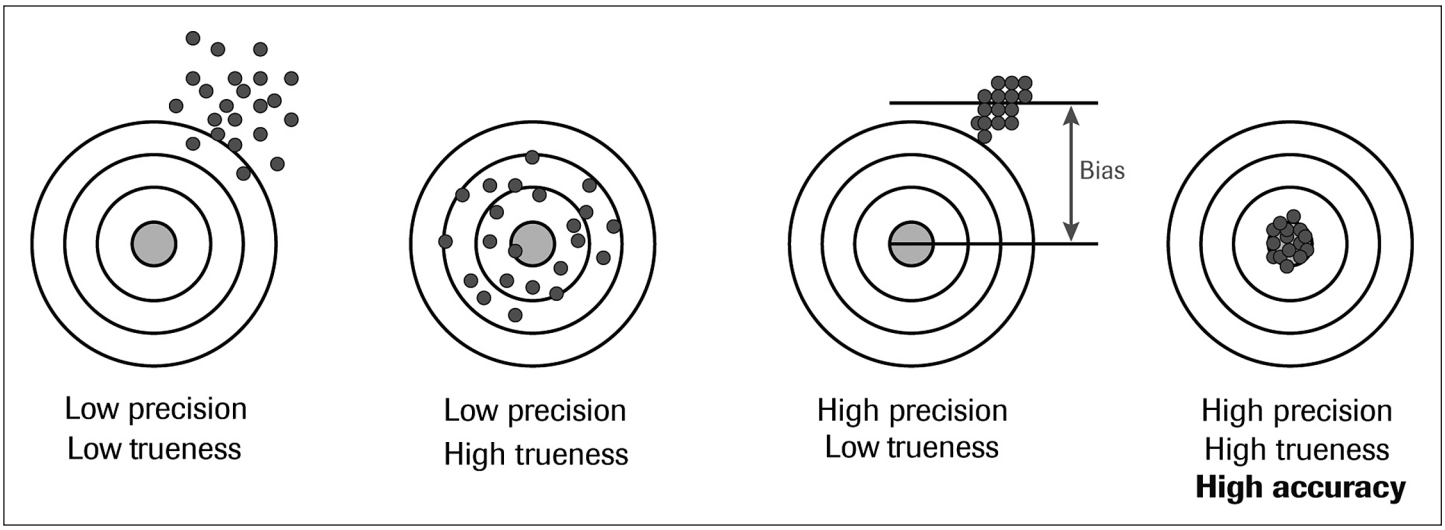


Figure 1. Precision, trueness, and accuracy of measurement.

Reference Method and Reference Measurement

Definitions of trueness and accuracy make reference to the term "accepted reference value." Use of the term "reference" or "reference method" causes a lot of confusion since this term has very different meanings in clinical chemistry and everyday language: while in clinical chemistry a reference method (or reference measurement procedure) usually refers

to a very specific, often rather sophisticated method with a well-known measurement error, in everyday language, a reference method is simply the method with which measurement results (e.g., of a BG system) are compared. For the latter case, the term "designated comparison method" is a better description and should therefore be preferred. In the context of clinical chemistry, an example for a reference method for glucose is isotope dilution gas chromatography mass spectrometry (ID-GC/MS). This is currently the best (i.e., most accurate) reference method for glucose. It is obvious that the analytical performance of the method with which measurement results are compared is of utmost importance.

As ID-GC/MS is a very complex and time-consuming method that can only be handled by a few laboratories worldwide, manufacturers usually choose methods[19] that are easier to handle to calibrate their BG systems. These so-called "manufacturer's standing measurement procedures" should, however, preferably be calibrated by ID-GC/MS to guarantee an unbroken traceability chain. Roche, for example, uses a hexokinase method[20] for whole blood as the manufacturer's standing measurement procedure: after a deproteinization step with perchloric acid, the samples are measured on a cobas 6000 analytical system (Roche Diagnostics, Rotkreuz, Switzerland). This method is also an internationally accepted reference method—although of lower metrological order than ID-GC/MS; however, it is directly calibrated to ID-GC/MS. The complete traceability chain for this case is shown in **Figure 2**.

Another method to which some manufacturers are calibrating their BG systems is the glucose oxidase method implemented on the YSI 2300 STAT Plus analytical system (Yellow Springs Instruments, Yellow Springs, OH), which, to our knowledge, has no ID-GC/MS in its traceability chain.

Fundamentally, the method used for the reference measurement should be clarified. According to the ISO 15197 standard, a detailed description of the manufacturer's standing measurement procedure used to determine the comparison values is required. It is a widespread misconception that "just any laboratory method" can be used as a reference.

Traceability, Uncertainty, Calibration, and Matrix Effect

The European Commission's *in vitro* diagnostics directive requires that "the traceability of values assigned to calibrators and/or control materials must be assured through available reference measurement procedures and/or available reference materials of a higher order."[21] The concept of traceability relates measurement values to a reference standard. It is defined as the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the so-called measurement uncertainty (characterizing the dispersion of the quantity values being attributed to the substance to be measured, in our case, glucose).[16]

The concept of measurement uncertainty[22] is different and considered complementary to the concept of total error by most authors. The total error concept can be considered a "bottom-up" approach that starts with the patient sample and tries to estimate its deviation from the "true value" by assessing different error components (usually random error and systematic error). On the contrary, the concept of measurement uncertainty is a "top-down" approach. It starts with the question, "How exact is the knowledge about my highest standard, e.g., the reference material (the "top")?" The concentration of this material already has some uncertainty, because a chemical composition can never be determined without uncertainty; the glucose content (or purity) of the glucose reference material SRM 917 (discussed later) is 99.7% ± 0.3%). When this material is dissolved to prepare a calibration solution of defined concentration to calibrate the reference method, further uncertainty is introduced because of the weighing procedure and the volume determination in this process. When the reference method is being used to calibrate in a next step "intermediate methods" and finally routine methods, further uncertainties need to be considered as well, and the uncertainty of the entire process becomes larger and larger. The uncertainties of each individual step are estimated separately and then added. By adding individual uncertainties of the individual uncertainties that are introduced at each level of the traceability chain, an estimate of the uncertainty of the result of the measurement of the patient sample can finally be obtained. Procedures how to determine measurement uncertainty are described in the publication *Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement (GUM)*.[23] According to the *In Vitro* Diagnostics Directive[21], manufacturers are obliged to provide information about the measurement uncertainty.

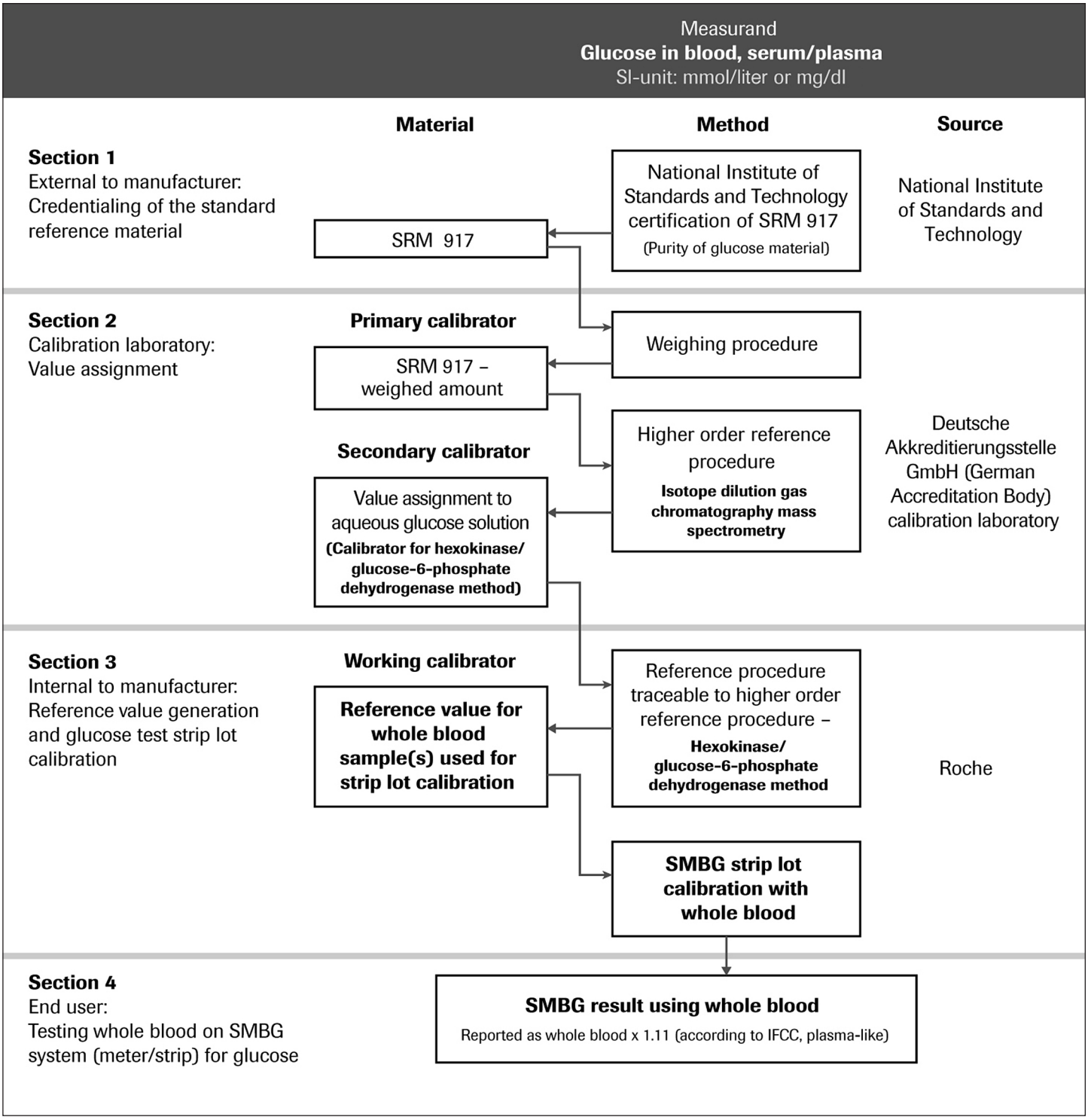


Figure 2. Example of a traceability chain (here depicted for Roche BG systems). IFCC, International Federation of Clinical Chemistry and Laboratory Medicine.

The aim of traceability is to link measurement results from a patient sample to a commonly accepted reference, making them internationally comparable across measurement systems, location, and time.[24] Since uncertainty increases cumulatively at each step in the traceability chain, it is advisable to omit as many steps as possible.[25] In more simple words, traceability means that through a set of suitable calibrators, BG systems are finally calibrated to a reference standard and that information about the measurement uncertainty is available.

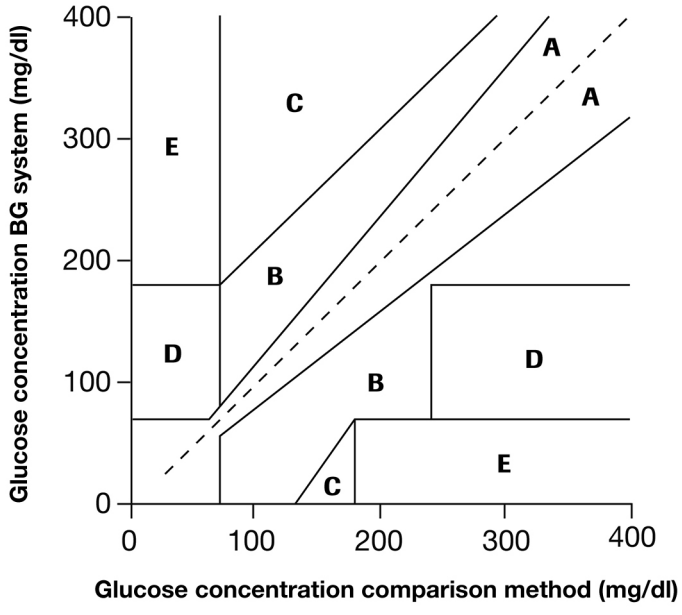
The highest standard available for glucose is a reference material by the U.S.-based National Institute of Standards, which is called SRM 917. Intuitively, one might think that it is sufficient to calibrate a glucose method with (an aqueous solution of) SRM 917 in order to establish perfect traceability. In reality, this is much more complicated. The reason is that a measurement result is determined not only by the substance we want to measure (e.g., glucose), but also by the matrix in which it is measured (e.g., whole blood, plasma, water). This so-called matrix effect needs to be considered and must be mathematically and experimentally compensated. For example, when glucose samples are measured with the ID-GC/MS reference procedure, the protein content of these samples needs to be carefully removed so that the glucose sample has the same (aqueous) matrix as the reference material that is used to calibrate the ID-GC/MS.

As mentioned, ID-GC/MS is a complex and time-consuming method that can be handled by only a few laboratories. Therefore, manufacturers often calibrate BG systems to a measurement procedure that is based on (sometimes modified) clinical laboratory methods.[19] Three different enzymatic reactions are utilized predominantly: glucose oxidase, hexokinase, and glucose dehydrogenase.[19,26] Hexokinase is reported to be less prone to analytical deviations by some authors.[26,27] However, for each and every one of these enzymes, a number of different coenzymes and mediator systems exist that influence the accuracy of BG measurement.[28]

Clinical Accuracy/Error Grid Analysis

Clinical accuracy of SMBG devices is usually determined by a so-called error grid analysis.[20,29] For an error grid analysis according to Clarke and coauthors,[30] all data pairs of BG measurements versus the comparison method are noted in a graph that is divided into five zones describing clinical accuracy (**Figure 3**).[29] Subsequently, the number of data pairs in each zone is counted to provide single numbers to describe the clinical accuracy.

The importance of an analysis tool for the description of clinical accuracy is revealed through the following example of a BG value of 76 mg/dl, representing an impending hypoglycemia. According to the ISO 15197:2003 standards,[31] SMBG values in the range of 51 to 91 mg/dl would be considered accurate, though these two values at the end of the range would result in quite different therapeutic consequences.[29] ISO 15197:2003 loses its validity after a transitional period of 3 years. In May 2013, the revised ISO 15197:2013 was published.[32]

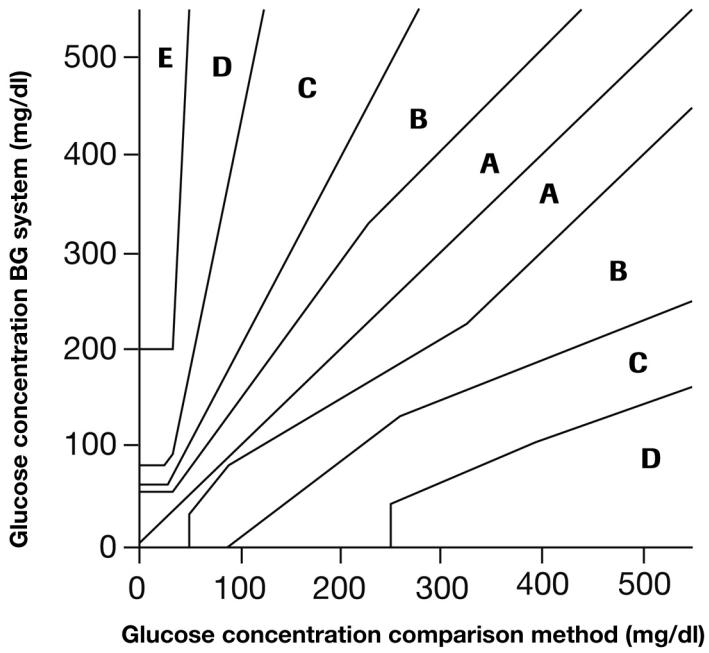


Zone	Description	Meaning
A	Clinically accurate, would lead to correct treatment decisions	Clinically acceptable
B	Would lead to benign decisions or no treatment	
C	Would lead to overcorrection of normal glucose levels	Clinically significant errors
D	Would lead to failure to detect and treat high or low glucose levels	
E	Would lead to erroneous treatment decisions	

Figure 3. The Clarke error grid.[20]

Error grid analysis also enables quantification of the clinical accuracy of SMBG measurements and describes the impact of the accuracy of BG systems on clinical decision making.[29,30,33,34]

Authors on the matter cite as problematic the fact that the A zone (clinically acceptable result) is contiguous to a D zone (clinically significant error).[35] Thus, two results with almost the same amount of error could be assigned to very different clinical outcomes in the Clarke error grid. This problem is resolved by the Parkes error grid,[33] which follows a somewhat different approach that is described in **Figure 4**. The Clarke and Parkes error grids have different demarcations for zones A to E. In the Parkes error grid, zone B only borders A and C, zone C only borders B and D, and so on.[35] Use of the Parkes error grid is proposed by the revised ISO 15197:2013; however, it is not routinely used thus far.[32]



Zone	Description
A	No effect on clinical action
B	Altered clinical action; little or no effect on clinical outcome
C	Altered clinical action; likely to affect clinical outcome
D	Altered clinical action; could have significant medical risk
E	Altered clinical action; could have dangerous consequences

Figure 4. The Parkes error grid.[33]

Assessment of Blood Glucose Systems

Self-monitoring of blood glucose performance should support patients in optimizing their glycemic control.[36] This can only be achieved when the measured BG values are a reliable source of information for appropriate and immediate therapeutic decisions. Thus, as stated earlier, the BG system's accuracy is an imperative aspect for the reliability of results.[37] Reliability is defined as the probability that a given item will perform its intended function for a given period of time under a given set of conditions.[38] Therefore, the question arises as to how the accuracy of BG systems should be evaluated and described adequately.

Evaluation Principles According to ISO 15197

Requirements for BG systems, including accuracy, are prescribed in detail in the internationally accepted standard EN ISO 15197: 2003.[31] According to this ISO standard, "system accuracy" shall be evaluated with capillary blood samples collected from at least 100 different subjects (with BG concentrations ranging from ≤50 to >400 mg/dl) over at least 10 days. The evaluation shall be performed with at least 100 fresh samples, each with sufficient volume to be measured by two different BG systems and at least in duplicate by the manufacturer's standing measurement procedure (which, as explained earlier, is usually different for BG systems from different manufacturers).[31]

This requirement is important and often neglected in publications of instrument evaluations. Investigators often choose just "their laboratory method," as the manufacturer's standing measurement procedure is not available to them.

It is also important to compare "like with like," i.e., capillary blood samples must not be compared with plasma samples, as the glucose concentrations can be very different in both compartments, in particular, in the postprandial state.[39]

According to ISO 15197:2013, 95% of the BG results shall fall within ±15 mg/dl of the reference method at BG concentrations <100 mg/dl and within ±15% at BG concentrations ≥100 mg/dl.[32] Ninety-nine percent of the results shall be located in the A (no effect on clinical action) and B (altered clinical action or little or no effect on clinical outcome) zones of the Parkes error grid.[32] **Figure 5** demonstrates the impact of tightened accuracy requirements on the number of results outside the accepted range.

Sample Material

In general, capillary or venous blood is used for glucose measurement. Unfortunately, the terms plasma or capillary whole blood calibration are sometimes mixed up, hence the consequent risk for misinterpretation.[40] Therefore, the International Federation of Clinical Chemistry and Laboratory Medicine recommends reporting all results only as glucose concentration in plasma:[39] a constant factor of 1.11 for conversion between the glucose concentration in capillary whole blood and the equivalent concentration in capillary plasma has been recommended for a hematocrit of 43%. For deviating hematocrit values, further mathematical corrections are necessary.[40] In the meantime, most manufacturers provide BG systems that follow this recommendation.

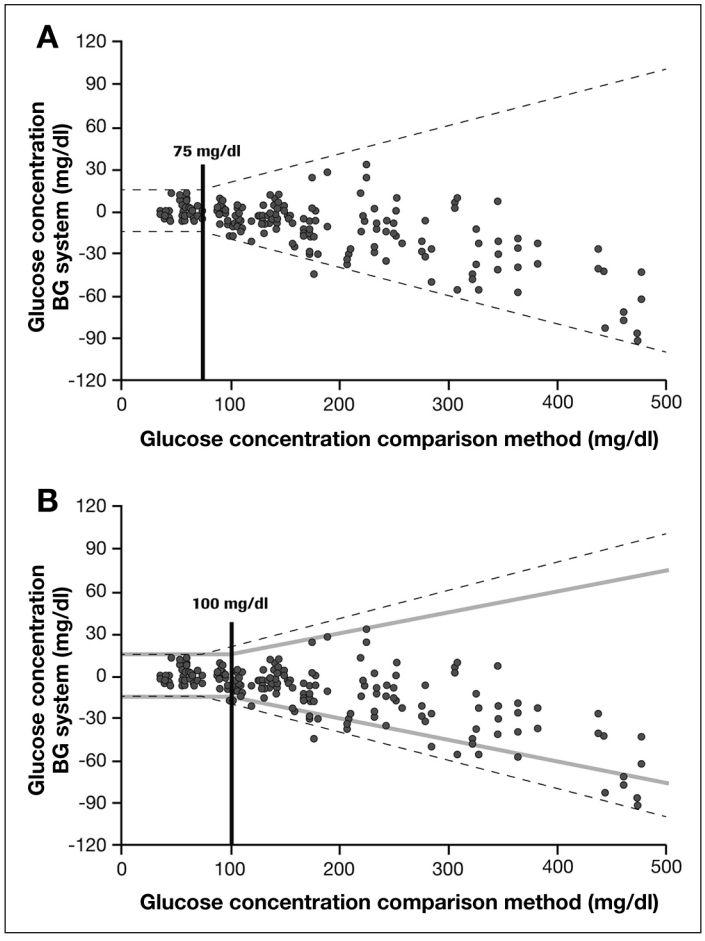


Figure 5. Accepted BG results according to **(A)** ISO 15197:2003 and **(B)** ISO 15197:2013. Requirements of the revised ISO 15197 provided, more results fall outside the accepted range.

Conclusion

Reliability of BG measurement is a prerequisite for successful diabetes self-management. It is essential to provide adequate patient education and optimize the analytical performance of BG systems. In this regard, consistent definition of key terms such as accuracy, trueness, precision, traceability, calibration, and matrix effect avoids misunderstandings and blurring. Definitions need to be based on metrological literature and international norms. The use of a uniform and standardized terminology by all partners is highly recommended.

The tightened standards of the revised ISO 15197:2013 should now be applied, although ISO 15197:2003 may still be used for a transitional period of 3 years. Currently, ID-GC/MS is the best available (i.e., most accurate) reference method for glucose. However, ID-GC/MS is a very complex method that can be performed only by a few laboratories worldwide. Therefore, manufacturers usually choose alternative methods. These "manufacturer's standing measurement procedures" should, however, preferably be calibrated by ID-GC/MS to guarantee an unbroken traceability chain. Evaluation of BG systems pursuant to the ISO standards, combined with the use of standardized terminology should be a precondition for future research on BG measurement.

Funding:

The creation of the manuscript was supported by an unrestricted educational grant of Roche Diagnostics GmbH, Mannheim, Germany.

Acknowledgments:

We would like to thank Lutz Heinemann, Guido Freckmann, Bernd Kulzer, and Volker Lodwig for their valuable input. Scientific writing was supported by Roche Diagnostics GmbH, Mannheim, Germany.

References:

1. Abraira C, Colwell JA, Nuttall FQ, Sawin CT, Nagel NJ, Comstock JP, Emanuele NV, Levin SR, Henderson W, Lee HS. Veterans Affairs Cooperative Study on glycemic control and complications in type II diabetes (VA CSDM). Results of the feasibility trial. Veterans Affairs Cooperative Study in Type II Diabetes. Diabetes Care. 1995;18(8):1113–23.
2. Durán A, Martín P, Runkle I, Pérez N, Abad R, Fernández M, Del Valle L, Sanz MF, Calle-Pascual AL. Benefits of self-monitoring blood glucose in the management of new-onset type 2 diabetes mellitus: the St Carlos Study, a prospective randomized clinic-based interventional study with parallel groups. J Diabetes. 2010;2(3):203–11.
3. Fisher L, Polonsky WH, Parkin CG, Jelsovsky Z, Petersen B, Wagner RS. The impact of structured blood glucose testing on attitudes toward self-management among poorly controlled, insulin-naive patients with type 2 diabetes. Diabetes Res Clin Pract. 2012;96(2):149–55.
4. Franciosi M, Lucisano G, Pellegrini F, Cantarello A, Consoli A, Cucco L, Ghidelli R, Sartore G, Sciangula L, Nicolucci A; ROSES Study Group. ROSES: role of self-monitoring of blood glucose and intensive education in patients with Type 2 diabetes not receiving insulin. A pilot randomized clinical trial. Diabet Med. 2011;28(7):789–96.
5. International Diabetes Federation. IDF guideline on self-monitoring of blood glucose in non-insulin treated type 2 diabetes. *http://www.idf.org/guidelines/self-monitoring*.
6. Kempf K, Kruse J, Martin S. ROSSO-in-Praxi follow-up: long-term effects of self-monitoring of blood glucose on weight, hemoglobin A1c, and quality of life in patients with type 2 diabetes mellitus. Diabetes Technol Ther. 2012;14(1):59–64.
7. Klonoff DC, Blonde L, Cembrowski G, Chacra AR, Charpentier G, Colagiuri S, Dailey G, Gabbay RA, Heinemann L, Kerr D, Nicolucci A, Polonsky W, Schnell O, Vigersky R, Yale JF; Coalition for Clinical Research-Self-Monitoring of Blood Glucose Scientific Board. Consensus report: the current role of self-monitoring of blood glucose in non-insulin-treated type 2 diabetes. J Diabetes Sci Technol. 2011;5(6):1529–48.
8. Polonsky WH, Fisher L, Schikman CH, Hinnen DA, Parkin CG, Jelsovsky Z, Petersen B, Schweitzer M, Wagner RS. Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. Diabetes Care. 2011;34(2):262–7.
9. Schnell O, Alawi H, Battelino T, Ceriello A, Diem P, Felton A, Grzeszczak W, Harno K, Kempler P, Satman I, Vergès B. Addressing schemes of self-monitoring of blood glucose in type 2 diabetes: a European perspective and expert recommendation. Diabetes Technol Ther. 2011;13(9):959–65.
10. Rydén L, Standl E, Bartnik M, Van den Berghe G, Betteridge J, de Boer MJ, Cosentino F, Jönsson B, Laakso M, Malmberg K, Priori S, Ostergren J, Tuomilehto J, Thrainsdottir I, Vanhorebeek I, Stramba-Badiale M, Lindgren P, Qiao Q, Priori SG, Blanc JJ, Budaj A, Camm J, Dean V, Deckers J, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo J, Zamorano JL, Deckers JW, Bertrand M, Charbonnel B, Erdmann E, Ferrannini E, Flyvbjerg A, Gohlke H, Juanatey JR, Graham I, Monteiro PF, Parhofer K, Pyörälä K, Raz I, Schernthaner G, Volpe M, Wood D; Task Force on Diabetes and Cardiovascular Diseases of the European Society of Cardiology (ESC); European Association for the Study of Diabetes (EASD). Guidelines on diabetes, pre-diabetes, and cardiovascular diseases: executive summary. The Task Force on Diabetes and Cardiovascular Diseases of the European Society of Cardiology (ESC) and of the European Association for the Study of Diabetes (EASD). Eur Heart J. 2007;28(1):88–136.
11. Schnell O AH, Battelino T, Ceriello A, Diem P, Felton A, Grzeszczak W, Harno K, Kempler P, Satman I, Vergès B. Consensus statement on self-monitoring of blood glucose in diabetes: a European perspective. Diab Metabol Heart. 2009;18(4):285-9.
12. Peel E, Parry O, Douglas M, Lawton J. Blood glucose self-monitoring in non-insulin-treated type 2 diabetes: a qualitative study of patients' perspectives. Br J Gen Pract. 2004;54(500):183–8.
13. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet. 1998;352(9131):837–53.
14. International Organization for Standardization. ISO 5725-1:1994: Accuracy (trueness and precision) of measurement methods and results.
15. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accred Qual Assur. 2007;12:45–7.
16. Joint Committee for Guides in Metrology. International vocabulary of metrology - Basic and general concepts and associated terms (VIM). 3rd ed. 2008. *www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf*. Accessed January 14, 2013.
17. Krouwer JS. Setting performance goals and evaluating total analytical error for diagnostic assays. Clin Chem. 2002;48(6 Pt 1):919–27.
18. Dybkaer R. Reference materials and reference measurement systems in laboratory medicine. Harmonization of nomenclature and definitions in reference measurement systems. Eur J Clin Chem Clin Biochem. 1995;33(12):995–8.
19. Tonyushkina K, Nichols JH. Glucose meters: a review of technical challenges to obtaining accurate results. J Diabetes Sci Technol. 2009;3(4):971–80.
20. Freckmann G, Baumstark A, Jendrike N, Zschornack E, Kocher S, Tshiananga J, Heister F, Haug C. System accuracy evaluation of 27 blood glucose monitoring systems according to DIN EN ISO 15197. Diabetes Technol Ther. 2010;12(3):221–31.
21. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices, Annex I A 3. Official Journal L 331, 07/12/1998.

22. Working Group 1 of the Joint Committee for Guides in Metrology. Evaluation of measurement data — guide to the expression of uncertainty in measurement. 2008. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed July 1, 2013.

23. Joint Committee for Guides in Metrology. Evaluation of measurement data — guide to the expression of uncertainty in measurement (GUM). 2008. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed September 6, 2013.

24. International Organization for Standardization. ISO 17511:2003: In vitro diagnostic medical devices -- measurement of quantities in biological samples -- metrological traceability of values assigned to calibrators and control materials.

25. Vesper HW, Thienpont LM. Traceability in laboratory medicine. Clin Chem. 2009;55(6):1067–75.

26. Weber C, Neeser K. Glucose information for tight glycemic control: different methods with different challenges. J Diabetes Sci Technol. 2010;4(5):1269–75.

27. Pelletier O, Arratoon C. Precision of glucose measurements in control sera by isotope dilution/mass spectrometry: proposed definitive method compared with a reference method. Clin Chem. 1987;33(8):1397–402.

28. Toghill KE, Compton RG. Electrochemical non-enzymatic glucose sensors: a perspective and an evaluation. Int J Electrochem Sci. 2010;5(9):1246–301.

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

30. 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.

31. International Organization for Standardization. ISO 15197:2003: *In vitro* diagnostic test systems -- requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus.

32. International Organization for Standardization. ISO 15197:2013: *In vitro* diagnostic test systems -- requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus.

33. 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.

34. Clarke W, Kovatchev B. Statistical tools to analyze continuous glucose monitor data. Diabetes Technol Ther. 2009;11 Suppl 1:S45–54.

35. 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.

36. American Diabetes Association. Standards of medical care in diabetes--2012. Diabetes Care. 2012;35 Suppl 1:S11–63.

37. Heinemann L, Lodwig V, Freckmann G. Accuracy in blood glucose measurement: what will a tightening of requirements yield? J Diabetes Sci Technol. 2012;6(2):435–43.

38. Wikipedia. Reliability engineering. http://en.wikipedia.org/wiki/Reliability_engineering. Accessed September 13, 2013

39. Walsh J, Roberts R, Vigersky RA, Schwartz F. New Criteria for Assessing the Accuracy of Blood Glucose Monitors meeting, October 28, 2011. J Diabetes Sci Technol. 2012;6(2):466–74.

40. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, Külpmann WR, Larsson L, Lewenstam A, Maas AH, Mager G, Naskalski JW, Okorodudu AO; International Federation of Clinical Chemistry Scientific Division Working Group on Selective Electrodes and Point of Care Testing. Approved IFCC recommendation on reporting results for blood glucose (abbreviated). Clin Chem. 2005;51(9):1573–6.