Accuracy of BG Meters and CGM Systems: Possible Influence Factors for the Glucose Prediction Based on Tissue Glucose Concentrations

Guido Freckmann, Stefan Pleus, Manuela Link and Cornelia Haug

Abstract The goal of this paper is to describe the metrics used for the evaluation of accuracy of blood glucose (BG) meters for self-monitoring of blood glucose (SMBG) and continuous-glucose monitoring (CGM) system and their limitations and to discuss the current status of SMBG and CGM accuracy. SMBG measurement is used by patients for therapy control and for calculation of appropriate insulin doses for approximately 30 years. The minimum accuracy criteria for SMBG meters are currently defined by ISO 15197:2003 (at least 95% of results within ±20% or ±15 mg/dL of the comparison method measurement results for BG concentrations above or below 75 mg/dL, respectively). In 2013, these accuracy limits were revised in the standard ISO 15197:2013: at least 95% of results within ±15% or ±15 mg/dL for BG above or below 100 mg/dL, respectively. SMBG systems are also used by patients for calibration of CGM systems. Therefore, precision and trueness of the SMBG system are influencing the accuracy of the CGM results. The timing of the BG measurement used for calibration has to be taken into account because, during rapid glucose changes, a time lag exists between BG and the tissue glucose that is measured by CGM systems. The accuracy of CGM devices is often reported by the mean absolute relative deviation (MARD) between CGM results and BG comparison results. This parameter is influenced by different factors like study procedures, glucose fluctuations during the study, and distribution of comparison BG measurements. It is important to define standard study procedures and evaluations to be able to compare MARD results from different studies. For the correct prediction of glucose concentrations, the specific prediction method as well as the accuracy of the CGM system, which may be affected by the accuracy of the SMBG system used for calibration, and the timing of the calibration are important aspects.
1 Introduction

Methods for prediction of future glucose concentration based on previous glucose using continuous subcutaneous glucose monitoring (CGM) data are currently under development. A prediction of future glucose could be beneficial to patients; early detection of glucose concentration falling below certain low glucose thresholds could, for example, help to prevent hypoglycemia. Glucose prediction can be combined with algorithms for closed-loop control to optimize glycemic control.

In current therapy, most patients treated with multiple daily insulin injections (MDI) or continuous subcutaneous insulin infusion (CSII) use self-monitoring of blood glucose (SMBG). SMBG measurement is used by patients for therapy control and for calculation of appropriate insulin doses for more than 30 years. MDI and CSII patients measure 3–8 single spot blood glucose values per day. For adequate glucose prediction, a much higher frequency of glucose values is needed. This higher frequency of glucose values can, for example, be achieved by using CGM systems.

In the last decade, different CGM systems became available that provide a glucose value every 1–10 min. The current CGM systems are measuring the glucose in the subcutaneous tissue and are calibrated against SMBG values. Thus, SMBG accuracy influences the accuracy of CGM systems and, consequently, the accuracy of glucose predictions based on CGM results. Additionally, the timing of the calibration is an important aspect, since time delays between BG and tissue glucose, which is measured by CGM results, can be observed when glucose concentration is changing rapidly.

The goal of this paper is to describe the metrics used for the evaluation of accuracy of SMBG meters and CGM systems and their limitations and to discuss the current status of SMBG and CGM accuracy.

2 SMBG Accuracy and CGM Calibration with SMBG Results

All currently available CGM systems are calibrated on blood glucose values provided by SMBG. The accuracy of the BG value and the timing of the BG measurement used for calibration are important for the accuracy of the CGM results.

2.1 SMBG Accuracy

SMBG measurement is currently used by patients for therapy control and for calculation of appropriate insulin doses. Therefore, the measured blood glucose values should be accurate to avoid miscalculation of insulin doses or failure to detect hypoglycemia. The American Diabetes Association stated in 1987, that a total
accuracy of 10% should be achieved [6]. In 1994, this goal was revised to an analytical accuracy of 5% [1]. Today, more than 20 years later, these accuracy goals are still not achieved by most SMBG meters.

The minimum acceptable accuracy for results produced by a SMBG system is currently defined by ISO 15197:2003: at least 95% of results have to be within ±20% or ±0.83 mmol/L (15 mg/dL) of the comparison method measurement results (manufacturers measurement procedure) for BG concentrations above or below <4.2 mmol/L (<75 mg/dL), respectively [15]. In addition, ISO 15197 calls for the display of the system accuracy results of SMBG systems for glucose concentrations <4.2 mmol/L (75 mg/dL), as the percentage of values falling within the following intervals: ±0.28 mmol/L (±5 mg/dL), ±0.56 mmol/L (±10 mg/dL), and ±0.83 mmol/L (±15 mg/dL). For glucose concentrations >4.2 mmol/L (75 mg/dL), results shall be expressed as the percentage of values falling within the following intervals: ±5, ±10, ±15 and ±20%.

In 2013, this ISO standard was revised (ISO 15197:2013) with the following accuracy criteria: 95% of results have to be within ±15% or ±15 mg/dL for BG concentrations above or below 100 mg/dL, respectively [16]. The new ISO standard also asks for the stricter limits in evaluations where the BG meter is used by laypersons. It also requires the assessment of clinical accuracy using the consensus error grid [16, 25]. Only limited data is available showing the accuracy in the hands of users, especially according to the revised ISO 15197. In Table 1, an example of how the system accuracy results can be displayed as recommended in ISO 15197:2013 is shown.

The difference plot in Fig. 1 is an example that shows the system accuracy results of three reagents lots of an SMBG system that was evaluated following ISO 15197 with all results within the limits of ISO 15197:2003 and 15197:2013 [29]. Unfortunately, the performance of the SMBG devices often is lower than shown in the example.

| Table 1 | Example for presentation of system accuracy results as recommended in ISO 15197:2013 |
|-------------------------------------------------------------|
| System accuracy results for glucose concentration <5.55 mmol/L (<100 mg/dL) |
| Within ±0.28 mmol/L (Within ±5 mg/dL) | Within ±0.56 mmol/L (Within ±10 mg/dL) | Within ±0.83 mmol/L (Within ±15 mg/dL) |
| 68/150 (45.3%) | 105/150 (70.0%) | 143/150 (95.3%) |

| System accuracy results for glucose concentration ≥ 5.55 mmol/L (≥ 100 mg/dL) |
| Within ±5% | Within ±10% | Within ±15% |
| 221/450 (49.1%) | 383/450 (85.1%) | 439/450 (97.6%) |

| System accuracy results for glucose concentrations between X.XX mmol/L (XX mg/dL) and YY.Y mmol/L (YYY mg/dL) |
| Within ±0.83 mmol/L or ±15% (Within ±15 mg/dL or ±15%) |
| 582/600 (97.0%) |

Note X.XX mmol/L (XX mg/dL) and YY.Y mmol/L (YYY mg/dL) stand for the lowest and highest glucose concentration measured with the comparison method, respectively.
Multiple studies were published during the last few years, testing meters applying the limits of the ISO 15197:2003 and ISO 15197:2013 standards. [3, 4, 9, 11, 12, 20, 26]. The quality of such studies differs, as only some studies are following the ISO 15197:2003 study protocol, while others have major deviations from the protocol [31]. All of these studies were comparing the system accuracy when applying the ISO 15197:2003 limits and they found that between 60 and 100% of the investigated SMBG systems have at least 95% of results within the ISO 15197:2003 limits. In some of these studies, the ISO 15197:2013 limits were applied as well, which were fulfilled by less than half of the investigated test strip lots. The observed differences in system accuracy can be attributed to or is influenced by a list of reasons including, but not limited to, the production process, the type and quality of test strip coding, user handling [30], and the manufacturers measurement method used for calibration. Measurement methods are reported to have systematic differences of up to 8% [32]. Such factors can lead to systematic or random measurement errors.

2.2 CGM Calibration with SMBG Results

Tissue glucose is not readily available for measurement, thus the more easily available BG is measured by patients with SMBG systems in order to be used for calibration of CGM devices. However, when calibrating against SMBG results, not only does the accuracy of the SMBG system influence the quality of the calibration, but also the timing of the BG measurement subsequently used for CGM calibration.
The accuracy, i.e., precision and trueness, of the SMBG system used for calibration are influencing the accuracy of the CGM results. In 2009, Kamath and colleagues reported that the MARD of the CGM system they used could be nearly halved (16.0% vs. 8.5%) when switching from calibration against SMBG results to calibration against results from a laboratory method [17]. It is unclear whether the effect could be reduced with top-of-the-line SMBG systems, thus reducing the initial MARD.

Additionally, a time delay between the glucose values provided by BG meters and CGM systems is observed during rapid glucose changes. This time delay is composed by a physiologic time delay, which is independent from a specific CGM system used, and a technical time delay caused by the specific CGM system. The physiologic time delay reflects the time required for the diffusion of glucose from the blood capillaries into the subcutaneous tissue [18]. The technical time delay is caused by two main factors. First, the glucose has to diffuse through the CGM sensor membrane and onto the sensor; and second, the raw signal from the CGM sensors often is smoothened by an algorithm, which further increases the time delay. This time delay occurs at all times and is not limited to rapid glucose changes; however, it is most clearly visible during rapid glucose changes and its contribution to a measurement error is more pronounced during these rapid glucose changes than when glucose only changing slowly. Rapid changes of glucose are observed frequently after the ingestion of meals or during exercise. Subsequently, the timing of the calibration is important and CGM systems should be calibrated at times of minimal BG change. Zueger and colleagues found lower MARD when using preprandial calibration as compared to using postprandial calibration [35].

SMBG accuracy and CGM calibration are the two most common influences on the accuracy of CGM systems. However, there are other aspects as well. In one study, it was suspected that the SMBG system showed a pronounced variability between vials of test strips (see Fig.2). In other cases, contaminated hands were found to heavily affect the accuracy of an otherwise good SMBG system [13, 14]. However, these influence factors are not common and they can be avoided by taking appropriate steps.

In summary, calibration of CGM systems should only be performed with high-accuracy SMBG systems, and phases of rapid glucose changes, i.e., within 1–2 h after carbohydrate intake, should be avoided for calibration of CGM systems.

In this context it should be noted that, recently, a factory-calibrated tissue glucose monitoring system was introduced in the European market. The term factory-calibrated is used for systems that do not have to be calibrated while wearing, as they have an internal calibration implemented during the system’s production. It is possible that in the near future, more and more systems will be factory-calibrated. As such factory-calibrations can be performed in a controlled environment, it is highly likely that the error introduced by inappropriate calibration will be markedly lower than that introduced by online, prospective calibration with SMBG systems (Fig.2).
3 Accuracy of CGM Systems

In the literature, the accuracy of CGM systems is assessed by a number of different parameters. In the following section, two of these parameters are described in more detail, namely the mean absolute relative difference (MARD) and the precision absolute difference (PARD).

The MARD indicates how closely CGM results and BG results match. The PARD on the other hand indicates how closely two sensor traces of the same CGM system worn by the same subject at the same time follow each other.

A third parameter, the continuous-glucose error grid analysis (CG-EGA), while not described in detail, is worth mentioning. The CG-EGA provides a clinical assessment of the CGM systems accuracy. However, BG results have to be obtained at least every 15 min in order to adequately perform CG-EGA, which places an additional burden on both the study participants and the study personnel.

3.1 Mean Absolute Relative Difference

As stated above, the MARD indicates how closely CGM results and BG results match. The MARD is commonly used to assess CGM accuracy, and it is calculated
as the average of the absolute values of the individual relative differences between CGM results and BG results (see Eq. 1).

\[
MARD = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| CGM_i - BG_i \right|}{BG_i} \times 100\% \tag{1}
\]

In Eq. 1, \(BG_i\) is the result of the \(i\)th BG measurement, \(CGM_i\) is the corresponding CGM result and \(N\) is the total number of pairs of BG results and CGM results.

Despite being commonly used, MARD results from different studies may not necessarily be comparable, since the MARD is influenced by a number of factors [24]: The MARD can be affected by the selection of the study participants, since their specific glucose values can influence MARD, and the study protocol, namely the rates of change (because of the time delay at rapid glucose changes, especially if glucose excursions are induced) and the duration of hypo- and hyperglycemic episodes. Nielsen and colleagues, for example, found that the MARD if the same CGM system was lower in type-2 diabetes subjects than in type-1 diabetes subjects, who are known to show higher glycemic variability [23].

Since these influence factors may make comparisons between different studies difficult, it is worth mentioning that the Clinical and Laboratory Standards Institute (CLSI) has issued a guideline, in which some specifications toward study protocols and data analysis are made [5]. In 2013, data from two clinical trials were published, in which the performance of, in total, five CGM systems was assessed [10, 28, 34]. Three CGM systems were used in the same study, whereas the other two systems were used in another study with very similar study procedures, thus allowing for a comprehensive comparison of all five CGM systems.

As stated above, an SMBG system used for calibration should provide adequate analytical performance especially if SMBG is used for calculating the MARD values. In many studies, devices are calibrated using SMBG values, but the evaluation of MARD is based on values obtained by a laboratory method. This seems to be the superior method, but this may lead to additional error if there is a systematic measurement error between the laboratory method and the SMBG system. In one of the abovementioned comparison studies, a systematic measurement error of 14% between the SMBG system used for calibration and the laboratory method was found [10]. Since most evaluations in that study were made against SMBG data, this was not a problem, but it may have lead to problems for evaluations in which CGM data were compared to measurement data from the laboratory method.

While the MARD is a parameter that is easy to compute and to interpret, the causes of low performance (i.e., high MARD) cannot be identified in detail. While a low MARD indicates close approximation of the CGM results to the BG results, higher MARD values give no indication about systematic measurement error or its sign (positive or negative error) or random measurement error. Figure 3 shows an example of a CGM trace and the corresponding BG measurements.

CGM systems that are currently available or were available until recently, show MARD results between approximately 10 and 20% [7, 10, 28, 34].
Fig. 3  Example of a CGM trace (blue line) and BG measurements (red diamonds) used for MARD calculation. Note that at approximately 07:00, the BG result is smaller than the CGM result and at approximately 13:00 the BG result is higher than the CGM result. In the calculation, both incidents contribute towards a higher MARD result, i.e., lower performance, independent from the direction of the deviation

3.2 Precision Absolute Relative Difference

The PARD was introduced in 2009 as percentage ARD by Zisser and colleagues [33]. While this parameter is not as widely used as the MARD, it has regularly been used since its introduction to assess sensor-to-sensor difference of CGM systems.

In order to calculate the PARD, the same subject has to wear at least two sensors of the same CGM system. It is then calculated as the average of the absolute difference between the two sensor signals divided by their mean value (see Eq. 2).

\[
PARD = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{CGM_{i,1} - CGM_{i,2}}{(CGM_{i,1} + CGM_{i,2})/2} \right| \times 100\% \tag{2}
\]

In Eq. 2, \( CGM_{i,1} \) is the \( i \)th result of the first CGM sensor, \( CGM_{i,2} \) is the \( i \)th result of the second CGM sensor and \( N \) is the total number of CGM result pairs.

The major advantage of the PARD is that it ideally includes all CGM results obtained during the study, whereas the MARD is limited to those CGM results that have corresponding BG results. For CGM systems, which store one result per 1–10 min, this means that if there is no data loss, between 144 and 1440 results per day can be used in the assessment.

As with the MARD, the PARD is easy to compute and to interpret, but again, it lacks in detailed information. While low PARD results indicate that the sensor signals follow each other closely, high PARD results give no indication about the systematic or random measurement errors. Figure 4 shows two CGM traces obtained from sensors of the same CGM system worn by the same subject. CGM systems that are currently available or that were available until recently show PARD results between approximately 7 and 18% [2, 10, 28, 33, 34].
Accuracy of BG Meters and CGM Systems: Possible Influence Factors … 39

Fig. 4 Traces from two sensors of the same CGM system in the same subject. Note that between 00:00 and approximately 03:00, the Sensor B results are lower than the Sensor A results and between approximately 09:00 and 13:00 the Sensor B result is higher than the Sensor A result. Independent of the direction of this deviation, both incidents contribute to a higher PARD, i.e., lower performance.

4 Glucose Prediction Based on Tissue Glucose Concentrations

Glucose prediction requires adequate accuracy of CGM systems. This does not only include the commonly used MARD, but also the PARD. Ideally, CGM systems exhibit both low MARD numbers (i.e., CGM results and BG results match closely) as well as low PARD numbers (i.e., the CGM results from different sensors of the same CGM system follow each other closely). It is quite obvious that adequate agreement between CGM results and BG results and minimal time lag between the tissue and blood glucose concentrations are required for predicting the future BG from CGM results. On the other hand, the differences between CGM results of one sensor and another sensor of the same type should be sufficiently small, in order to allow for reproducible glucose prediction.

Glucose prediction has to take into account not only the accuracy of the CGM system itself, but also the accuracy of the measurement system that is used for calibration and the timing of the calibration.

Current CGM systems seem sufficiently accurate to allow for adequate glucose prediction with currently available prediction models [8, 19, 21, 22, 27].
References


Prediction Methods for Blood Glucose Concentration
Design, Use and Evaluation
Kirchsteiger, H.; Jørgensen, J.B.; Renard, E.; del Re, L.
(Eds.)
2016, XIV, 265 p. 93 illus., 72 illus. in color., Hardcover
ISBN: 978-3-319-25911-6