Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA): A Novel Approach for the Statistical Accuracy Assessment of Continuous Glucose Monitoring Systems

Manuel Eichenlaub, PhD[1]; Peter Stephan, MSc[2]; Delia Waldenmaier, PhD[1]; Stefan Pleus, PhD[1]; Martina Rothenbühler, PhD[3]; Cornelia Haug, MD[1]; Rolf Hinzmann, MD, PhD[4,8]; Andreas Thomas, PhD[5,8]; Johan Jendle, MD, PhD[6,8]; Peter Diem, MD[7,8]; Guido Freckmann, MD[1,8]

1 Institut für Diabetes-Technologie, Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm, Ulm, Germany

2 Mannheim, Germany

3 Diabetes Center Berne, Bern, Switzerland

4 Roche Diabetes Care GmbH, Mannheim, Germany

5 Pirna, Germany

6 Department of Medical Sciences, Örebro University, Örebro, Sweden

7 Endokrinologie Diabetologie Bern, Bern, Switzerland

8 IFCC Scientific Division - Working Group on Continuous Glucose Monitoring

Running title

Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA)

Corresponding Author

Manuel Eichenlaub, PhD, Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm, Lise-Meitner-Strasse 8/2, D-89081 Ulm, Germany, e-mail: manuel.eichenlaub@idt-ulm.de, Phone +49-731-50-990-0

Key words

accuracy, continuous glucose monitoring, FDA iCGM requirements, deviation intervals, sensor-to-sensor variability

Abstract

Background

The accuracy of continuous glucose monitoring (CGM) systems is crucial for the management of glucose levels in individuals with diabetes mellitus. However, the discussion of CGM accuracy is challenged by an abundance of parameters and assessment methods. The aim of this article is to introduce the Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA), a new approach for a comprehensive characterization of CGM point accuracy.

Methods

Data from two approved CGM systems was used to illustrate the CG-DIVA, which is based on the U.S. Food and Drug Administration requirements for “integrated” CGM systems.

Results

The CG-DIVA characterizes the expected range of deviations of the CGM system from a comparison method in different glucose concentration ranges using the concept of tolerance intervals, as well as the variability of accuracy within and between sensors. The results of the CG-DIVA are visualized in an intuitive and straightforward graphical presentation. Compared to conventional accuracy characterizations, the CG-DIVA infers the expected accuracy of a CGM system and highlights important differences between CGM systems. Furthermore, it provides information on incidence of large errors which are of particular clinical relevance. A software implementation of the CG-DIVA is freely available (https://github.com/IfDTUlm/CGM_Performance_Assessment).

Conclusions

We argue that the CG-DIVA simplifies the discussion and comparison of CGM accuracy and could replace the high number of conventional approaches. Future adaptations of the approach could thus become a putative standard for the accuracy characterization of CGM systems and serve as the basis for the definition of future CGM performance requirements.

Introduction

Continuous glucose monitoring (CGM) has been established as a valuable tool for managing glucose levels in individuals with diabetes mellitus. According to the Standards of Medical Care in Diabetes 2022, published by the American Diabetes Association (ADA), CGM should be offered for diabetes management in adults and youth on multiple daily injections or continuous subcutaneous insulin infusion.[1] Similarly, the European Association for the Study of Diabetes considers CGM technology as the standard for glucose monitoring in most adults with type 1 diabetes.[2] CGM systems are also an integral part of automated insulin delivery systems, which, according to the ADA, should be offered to youth and adults with type 1 diabetes.[1] Furthermore, glycemic goals for CGM-based metrics, such as the time in or below range, have been defined.[1–3] These developments highlight the importance of CGM system performance and reliability. One of the most crucial aspects of CGM performance is accuracy, which will be the focus of this article. In particular, it will deal with point accuracy, i.e., the closeness of agreement between CGM and corresponding data from a comparator device of higher measurement quality at a single point in time, as opposed to rate accuracy focusing on rate of change agreement.

Many different approaches to characterize CGM point accuracy have been utilized over the years. Despite ongoing criticism,[4–7] the mean absolute relative difference (MARD) between CGM and comparator measurements remains widely accepted. Another common parameter is the agreement rate (AR), which provides the percentage of CGM measurements within a certain range, e.g. ±20%, of their paired comparator measurements. An error grid analysis, originally developed for blood glucose monitoring systems (BGMS) but also applied to CGM systems, provides another parameter as well as a graphical visualization.[8–10] A dedicated error grid analysis for CGM systems that includes the rate of change in glucose levels has been proposed in 2004.[11] However, as its calculation and interpretation are considered to be very complex, it is rarely used today.[12,13] Tabulating the concurrence between CGM and comparator measurements in various glucose ranges has also been proposed.[14] Additionally, the calculation of correlation coefficients in the context of linear regression analysis is common.

In general, a single metric is not capable of conveying an adequate picture of CGM accuracy. To compensate for this and provide a more comprehensive characterization, MARD and AR are often stratified by glucose level, e.g. <70, 70-180 and >180 mg/dL (<3.9, 3.9-10.0 and >10.0 mmol/L). Furthermore, to provide information on the variability of accuracy between sensors, a histogram of sensor-specific MARD values is sometimes provided.

From this brief overview of parameters, it should become clear that the discussion of CGM accuracy is complicated by the abundance of existing approaches.[13] Therefore, there is a need to standardize the accuracy characterization and presentation for CGM devices to facilitate the comparison of different systems. In this context, the Clinical and Laboratory Standards Institute (CLSI) has published the POCT05 guideline on "performance metrics for continuous interstitial glucose monitoring".[14] However, the proposed approach for CGM point accuracy characterization names a number of different parameters and graphical displays including different MARDs, biases, ARs and error grid analyses, and therefore fails to alleviate the previously mentioned issue of abundance.

A more concise step towards the standardization of CGM accuracy characterization has been made in 2018 by the United States Food and Drug Administration (FDA) through the release of special control regulations, which, for the first time, included the definition of minimum acceptance requirements for "integrated" CGM systems.[15,16] In terms of point accuracy, the FDA requirements were mainly defined on the parameter of AR, in particular its lower one-sided 95% confidence bound, thus reducing the primary outcome to a single parameter. A more detailed discussion of the FDA requirements will be provided in the next section.

The aim of this article is to introduce the Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA), a novel approach for the comprehensive description of CGM point accuracy. It is based on the FDA requirements but provides additional characteristics that cover all relevant aspects of CGM point accuracy while allowing a straightforward evaluation and interpretation. It should thus appeal to manufacturers and clinical practitioners alike.

The present article supports the activities of a working group on CGM created by the Scientific Division of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC).[17]

Methods

FDA requirements

As the CG-DIVA is based on the FDA requirements for point accuracy (Table 1), they are briefly discussed here. Table 1 shows that the FDA requirements are defined on the lower bound of the confidence interval of the AR, instead of using AR directly, thus accounting for the statistical uncertainty of the AR. However, the method for calculating the confidence bounds is not specified, which could lead to different results and therefore compliance with the requirements depending on the method chosen. We will address this particular issue in a separate article.

Another aspect to highlight is the fact that, according to the FDA requirements, the glucose range is stratified by the glucose values measured by the CGM system, not by the comparator measurement method. This stratification with respect to the examined CGM system, not the comparator method, is highly unusual in the field of clinical chemistry and inconsistent to the FDA requirements for BGMS, for both over-the-counter and point-of-care use.[18,19] In both requirements for BGMS, it is recommended to present a difference plot, where the results of the comparator method are plotted against the difference between BGMS and comparator results. This procedure implies that deviations should be assessed with respect to the comparator results as they are assumed to be of higher measurement quality. Using CGM data for glucose range stratification means that the method for accuracy assessment depends on the accuracy of the CGM system, while it should only depend on the accuracy of the comparator method. As a result, the AR, e.g. <70 mg/dL (<3.9 mmol/L), does not directly characterize the accuracy during hypoglycemia, but rather when the CGM system indicates hypoglycemia, which, depending of the system performance could be quite different from true hypoglycemia. It thus also hinders the comparison of different CGM systems amongst each other.

More generally, the focus in ARs alone to characterize point accuracy can be challenged as they cannot provide direct information on bias and imprecision. Furthermore, a specific characterization of the variability in accuracy between individual sensors was not demanded by the FDA.

Data description and conventional accuracy characterization

The data used to exemplify the CG-DIVA was collected from 24 subjects and two approved CGM devices examined in the same clinical study. CGM system A required manual calibrations, whereas system B was factory-calibrated. The study had a duration of eight calendar days and the subjects wore both devices simultaneously. Comparator measurements were collected from capillary samples using a commercially available BGMS that was also used to calibrate CGM system A according to manufacturer's instructions. A total of 3989 and 3619 CGM-comparator measurement pairs were available for CGM system A and B, respectively, with approximately 7%, 63% and 30% of comparator measurements falling <70, 70-180 and >180 mg/dL (<3.9, 3.9-10.0 and >10.0 mmol/L), respectively.

To compare the novel approach with conventional accuracy parameters, ARs and their lower 95% confidence limits as well as MARDs were calculated.

Results

Conventional accuracy characterization

The results of a conventional accuracy characterization including ARs and MARDs are provided in Table 2. Furthermore, a histogram of MARD results is provided in Figure 1.

Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA) and graphical presentation

As mentioned before, the CG-DIVA presented in this article is based on the FDA requirements in Table 1. However, to avoid the disadvantages of stratifying the glucose ranges according to the CGM values, the CG-DIVA divides the glucose range according to the comparator measurement results, while keeping the same stratification into hypo-, normo- and hyperglycemia (<70, 70-180, >180 mg/dL (<3.9, 3.9-10.0, >10.0 mmol/L)). Furthermore, instead of ARs alone, the focus is put on deviations to provide direct information on bias and precision of the CGM system, which could be clinically more relevant than AR alone. In accordance with the FDA requirements, absolute deviations are calculated for comparator glucose levels <70 mg/dL (<3.9 mmol/L) and relative deviations for all others. Furthermore, a prerequisite for the application of FDA requirements as well as the CG-DIVA is the availability of a robust dataset obtained from a clinical study designed to fully represent the performance of the CGM system across the intended use population and measuring range.

The CG-DIVA itself is divided into the following two parts and its results are displayed in two separate plots described below. The results of applying it to the previously described datasets of CGM systems A and B are displayed in Figure 2. A table detailing the numerical results of the deviation intervals is provided in the supplementary materials.

Deviation intervals

The first part of the CG-DIVA infers intervals of deviations that can be expected from the CGM system in general. It thereby incorporates the statistical uncertainty of the data and thus follows the idea of the FDA of using confidence intervals for the definition of minimum requirements. For each comparator glucose range, the median deviation and the interval in which a certain central proportion of deviations are expected to be found are calculated, e.g.

85% and 98% for comparator glucose levels <70 mg/dL (<3.9 mmol/L). For the initial version of the CG-DIVA presented in this article, the size of these intervals was set to the AR requirements defined by the FDA (Table 1). The results of this part of the CG-DIVA, exemplified in Figure 2 (A) and (C), are displayed with a black line (median) and dark/light grey boxes (deviation intervals 1/2). The extent of the deviations is conveyed by the colored background based on limit 1 and 2 of the FDA requirements. The respective interval size of the boxes, based on the minimum AR requirements, is printed above the plot.

Examining the results of the CG-DIVA in Figure 2 (A) and (C) in relation to the corresponding ARs and their lower confidence bounds in Table 2, the following observations can be made. In both CGM systems, the deviation intervals for the total glucose range clearly extend into the red areas, which correctly indicates that corresponding lower AR confidence bounds are below the threshold of 87%. In general, the AR confidence bound requirements are likely not met whenever the corresponding deviation interval, i.e. both edges of the boxes, extends into the yellow and red areas, respectively. Conversely, if the deviation intervals lie within these areas, the requirements are most likely met. If only one interval edge is above the limit, e.g. for deviation interval 2 (light grey box) in the hyperglycemic range of system A, it is possible for the corresponding AR confidence bound (99.1%) to meet the respective requirement (99%). Here it should be mentioned that, in contrast to the FDA, the CG-DIVA uses the comparator glucose measurements to stratify the glucose range, meaning that the results cannot be directly interpreted in terms of compliance with FDA criteria.

Besides the context of AR, the deviation intervals allow the following interpretation of results. Taking the hyperglycemic range as an example, the size of deviation interval 1 (80%) supports the direct statement that one in five (20%) deviations are expected to fall outside the interval. As this interval represents the central portion of the expected deviations, it additionally follows that one in ten (10%) deviations can be found above its upper limit and one in ten (10%) deviations can be found below its lower limit. Using system A as an example (Figure 2 (A)), it can be stated that, for glucose levels >180 mg/dl (>10.0 mmol/L), one in ten (10%) deviations are expected to be larger than 20.3% and one in ten deviations are expected to be larger than -18.7%. Analogous statements about the limits beyond which one in two hundred (0.5%)

deviations occur, i.e. 37.2% and -48.0%, respectively, can be made from the size of deviation interval 2 (99%).

To infer the deviation intervals plotted with the boxes, the statistical concept of tolerance intervals was used. It is based on the well-known idea of describing the distribution of a dataset, in this case of deviations, by determining the range in which a given proportion of deviations are found, e.g. the central 90% range between the 5 and 95 percentiles. The tolerance interval provides a more universal proportion of deviations, e.g. 85%, that can generally be expected from the examined CGM system with a certain level of confidence, in this case 95%.[20]

For the calculation of the tolerance intervals a bias-corrected and accelerated bootstrap method[21,22] accounting for the clustered structure of the data was employed. This was necessary as the data recorded within individual sensors cannot be considered statistically independent. Instead, each CGM sensor forms a distinct cluster, which has to be preserved during the bootstrapping process. More details on the statistical calculations are provided in the supplementary materials. The same bootstrapping procedure was used to calculate the lower bound of the 95% confidence intervals of the ARs in Table 2.

Sensor-to-sensor variability

The second part of the CG-DIVA characterizes the variability in accuracy between and within individual sensors of the same CGM system, exemplified in Figure 2 (B) and (D). For that, each sensor is described by its median and central 90% range of deviations within every comparator glucose range. The sensors are ordered by their median deviations across the total glucose range. If less than ten data points are available for a sensor a determination of the 90% range is not sensible. In this case the full range of deviations is displayed and indicated by the capped bars. In contrast to the first part, it should be emphasized that this is a purely descriptive analysis of the collected data and without any statistical inference about the system.

Discussion

The CG-DIVA presented in this article focuses on the analysis of deviations between CGM and comparator measurements to characterize the accuracy of a CGM system. In particular the deviation intervals go beyond a basic description of the dataset and infer the system performance in general. For this, it is assumed that the dataset obtained from the CGM performance evaluation study is robust and somewhat representative of system's use in real-life. In comparison to ARs and MARDs, the deviation intervals can provide a more complete picture of CGM accuracy and reveal important differences between various systems. For example, the AR and MARD results of CGM systems A and B in the normoglycemic range, provided in Table 2, are very similar suggesting a comparable accuracy. However, comparing the corresponding results of the CG-DIVA from Figure 2 (A) and (C), it is revealed that system A shows a difference between approximately 5 and 10% in median and deviation interval limits compared to system B, which could affect clinical decision making. Inspecting the accuracy in the hypoglycemic range, the conventional parameters indicate better performance of system B, which is confirmed by the results of the CG-DIVA, especially through the decreased upper limit of deviation interval 1 in system B (25.0 vs. 39.9 mg/dL (1.4 vs. 2.2 mmol/L)). However, the CG-DIVA provides the additional information that system A has a considerable positive bias, thus increasing the risk of delayed or missed hypoglycemia treatment. Similar to previous works,[4–7] these results demonstrate that conventional accuracy parameters, especially the MARD, are not suitable to provide a comprehensive picture of the point accuracy of a CGM system.

Beyond bias, imprecision and AR, the CG-DIVA provides information on the incidence of particularly large deviations by allowing the statement that, e.g. one in ten deviations of CGM system A are expected to be larger than 39.9 mg/dL (2.2 mmol/L) in the hypoglycemic range. This information is clinically more meaningful than information on the average performance, because these large deviations can lead to severe errors in therapy decisions or automated insulin dosing. In fact, analyses of adverse event databases with respect to CGM inaccuracies have shown that large deviations occurring during CGM use outside clinical studies have been related to serious outcomes,[23,24] calling for a greater focus on large errors that occur only

sporadically. In this context, the AR might be an unsuitable parameter to characterize these types of errors because deviations outside their limits (±15 and ±40 mg/dL or %) are treated independent from their magnitude. This could be the rationale for the inclusion of requirement 3 by the FDA (Table 1). Furthermore, the AR treats positive and negative deviations equally, which might be inappropriate for the hypo- and hyperglycemic glucose ranges. In contrast, the CG-DIVA characterizes positive and negative deviations separately and is more sensitive to outliers thus providing a better characterization of large but sporadic deviations.

A further advantage of the CG-DIVA is that its results are presented in graphical form, which increases their interpretability and accessibility in comparison to ARs and MARDs usually presented in tabular form. Comparing the presentation of the deviation intervals to conventional graphical summaries such as difference or the error grid plots, the information is presented more concisely, allowing the identification of the most important accuracy characteristics and differences between CGM systems. The graphical presentation should therefore be accessible to people with a limited statistical background.

The second part of the CG-DIVA characterizes variability in accuracy from sensor to sensor, a crucial aspect of CGM performance that is often neglected. Examining the conventional approach of displaying a histogram of MARDs in Figure 1, it can be suspected that CGM system B has a higher sensor-to-sensor variability; however, the impact of this variability on the measurement results is not clear. In contrast, Figure 2 (B) and (D) give a more comprehensive characterization of sensor-to-sensor variability. Here, the larger variability between sensors in CGM system B is clearly shown by the differences in median sensor deviations across all comparator glucose ranges, which can be explained through the fact that system B is factory-calibrated. Furthermore, the results from the hypoglycemic range <70 mg/dL (<3.9 mmol/L) show that, despite a median bias close to zero, individual sensors of CGM system B can have a considerable positive bias, similar to system A. Additionally, it is demonstrated that, despite the larger variability in median deviations of system B, the imprecision of individual sensors, indicated by the antennae in the plot (90% range), is smaller in comparison to the manually calibrated system A.

Despite the described advantages of the CG-DIVA, there are several improvements to discuss. The stratification of the glucose range into hypo-, normo- and hyperglycemic ranges was based on the FDA requirements, which were in turn based on the clinical “Time in Range” limits.[3] While these ranges are well-established and can be used to characterize the status of glycemic control, they are perhaps unsuitable for assessing the accuracy of a CGM system. For example, hypoglycemia prevention measures like automatically suspending basal insulin infusion or raising a respective alert are typically made at glucose levels above 70 mg/dL (3.9 mmol/L), especially when falling. However, the glucose range relevant for hypoglycemia prevention, which might be between 70 and 100 mg/dL (3.9 and 5.55 mmol/L), is currently incorporated in the 70-180 mg/dL (3.9-10.0 mmol/L) range, disallowing a distinct assessment of accuracy in the crucial 70-100 mg/dL (3.9-5.55 mmol/L) range. It might thus be sensible to extend the first glucose range to glucose levels of up to 100 mg/dL (5.55 mmol/L). This would also solve the issue that there is a discontinuity in the quantification of deviations, when switching from absolute to relative deviations at glucose levels of 70 mg/dL (3.9 mmol/L). Here, it should be emphasized that the CG-DIVA can be adapted for any range size, given that a sufficient number of samples are available for computation of the expected range of deviations.

An additional improvement could be made by adjusting the deviation interval sizes displayed by the dark and light grey boxes. In its current form, the interval sizes are different for each comparator glucose range, thus making it difficult to compare the accuracy, in particular imprecision, across ranges. These interval sizes were chosen according to the FDA requirements for the ARs, but the method itself could be adapted to other interval sizes. For example, to allow clinically useful statements about deviations with an incidence of one in ten or one in one hundred, it might be pertinent to consistently choose the sizes of deviation intervals 1 and 2 to 80% and 98%, respectively. In this context, the design of the clinical study and especially the choice of sample size are important, as the incidence of rare events on the order of one in one hundred or less can only be reliably estimated with a sufficient number of data points in every comparator glucose range.

It should also be mentioned that the introduced methodology is, for the moment, only focused on point accuracy. However, CGM systems also provide information on the rate of change of glucose which is used for therapy decisions as well as automated insulin delivery. A reliable assessment of rate accuracy requires frequent comparator sampling over longer periods of time. Furthermore, the study design, in particular the initiation of rapid changes in glucose, affects the point accuracy. Future work will focus on extending the CG-DIVA with a similar characterization of rate accuracy. Additionally, an extension to include a description of sensor stability, i.e. the accuracy over the sensor use life, is planned.

To enable both a scientific and clinical discourse and to promote the use of CG-DIVA, a free and open-source software package allowing easy utilization of the approach on multiple platforms is published alongside this article (https://github.com/IfDTUlm/CGM_Performance_Assessment). Here, further details on the use of the software can be found.

Conclusions

The present article introduced the CG-DIVA, a novel approach for the comprehensive assessment of CGM point accuracy that focuses on the analysis of deviations, while also characterizing the variability in accuracy between sensors. By leveraging techniques from inferential statistics, the CG-DIVA provides universal information about the examined CGM system, beyond a simple description of the data collected in clinical studies. Furthermore, the approach summarizes the results in an easily accessible graphical form and incorporates all relevant aspects of CGM point accuracy, including the incidence of large deviations affecting clinical outcomes. We therefore argue that the CG-DIVA represents an attractive alternative to conventional point accuracy parameters such as MARD and AR, thus greatly simplifying the discussion and comparison of CGM point accuracy.

The first version of the CG-DIVA presented in this article was heavily influenced by the FDA requirements. However, several improvements and adaptations such as the harmonization of deviation interval sizes or the inclusion of a trend accuracy characterization have been discussed and will be subject of future publications. In the long term, we thus argue that further developments of the CG-DIVA could become the standard for the accuracy characterization and presentation of CGM systems, which is facilitated by the continuing publishing of freely accessible software implementations. Furthermore, it could serve as the basis for the definition of CGM performance standards, which is the main aim of the previously mentioned working group on CGM of the International Federation of Clinical Chemistry and Laboratory Medicine.

Acknowledgements

The authors would like to thank the members of the working group on continuous glucose monitoring created by the International Federation of Clinical Chemistry and Laboratory Medicine for their feedback and discussion regarding the content of this article. Furthermore, we would like to thank the Diabetes Center Berne for their financial support.

Authorship contribution

ME: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft

PS: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – review & editing

DW: Conceptualization, Methodology, Visualization, Writing – review & editing

SP: Methodology, Validation, Writing – review & editing

MR: Methodology, Validation, Writing – review & editing

CH: Writing – review & editing

AT: Writing – review & editing

JJ: Writing – review & editing

RH: Writing – review & editing

PD: Writing – review & editing

GF: Conceptualization, Writing – review & editing

Conflict-of-Interest Statement

GF is general manager and medical director of the IfDT (Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm, Ulm, Germany), which carries out clinical studies on the evaluation of BG meters, with CGM systems and medical devices for diabetes therapy on its own initiative and on behalf of various companies. GF/IfDT have received speakers' honoraria or consulting fees from Abbott, Ascensia, Berlin Chemie, Beurer, BOYDsense, CRF Health, Dexcom, i-SENS, Lilly, Metronom, MySugr, Novo Nordisk, PharmaSens, Roche, Sanofi, Sensile, Terumo and Ypsomed.

ME, DW, SP and CH are employees of the IfDT.

PS is an advisor to the IfDT

MR is an employee of Diabetes Center Berne.

RH is an employee of Roche Diabetes Care GmbH

JJ has received speakers' honoraria or consulting fees from Abbott, Ascensia, Eli Lilly, Medtronic, Novo Nordisk, and Sanofi.

AT is an independent consultant for diabetes technology. He was Scientific Director at Medtronic Diabetes Germany until 2020. He participated in advisory boards of the company Dexcom. He has received honoraria for lectures from Dexcom, Abbott and Berlin-Chemie (Menarini), among others.

PD is a board member of PharmaSens

Funding Source

This study was supported by Diabetes Center Berne, Switzerland.

References

1. American Diabetes Association Professional Practice Committee. 7. Diabetes Technology: Standards of Medical Care in Diabetes—2022. *Diabetes Care*. 2021;45(Supplement_1):S97-S112.

2. Holt RIG, DeVries JH, Hess-Fischl A, et al. The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia*. 2021;64(12):2609-2652.

3. Battelino T, Danne T, Bergenstal RM, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. *Diabetes Care*. 2019;42(8):1593-1603.

4. Kirchsteiger H, Heinemann L, Freckmann G, et al. Performance Comparison of CGM Systems: MARD Values Are Not Always a Reliable Indicator of CGM System Accuracy. *J Diabetes Sci Technol*. 2015;9(5):1030-1040.

5. Reiterer F, Polterauer P, Schoemaker M, et al. Significance and Reliability of MARD for the Accuracy of CGM Systems. *J Diabetes Sci Technol*. 2017;11(1):59-67.

6. Freckmann G, Pleus S, Grady M, Setford S, Levy B. Measures of Accuracy for Continuous Glucose Monitoring and Blood Glucose Monitoring Devices. *J Diabetes Sci Technol*. 2019;13(3):575-583.

7. Heinemann L, Schoemaker M, Schmelzeisen-Redecker G, et al. Benefits and Limitations of MARD as a Performance Parameter for Continuous Glucose Monitoring in the Interstitial Space. *J Diabetes Sci Technol*. 2020;14(1):135-150.

8. 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):622-628.

9. 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):1143-1148.

10. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014;8(4):658-672.

11. 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(8):1922-1928.

12. Wentholt IM, Hoekstra JB, DeVries JH. A Critical Appraisal of the Continuous Glucose–Error Grid Analysis. *Diabetes Care*. 2006;29(8):1805-1811.

13. Bailey TS. Clinical Implications of Accuracy Measurements of Continuous Glucose Sensors. *Diabetes Technol Ther*. 2017;19(S2):S-51.

14. Clinical and Laboratory Standards Institute (CLSI). *Performance Metrics for Continuous Interstitial Glucose Monitoring, 2nd Edition*. CLSI Guideline POCT05. Clinical and Laboratory Standards Institute; 2020.

15. U.S. Food and Drug Administration. EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR Dexcom G6 Continuous Glucose Monitoring System: Decision

Summary. Published online 2018. Accessed May 24, 2022. https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170088.pdf

16. Garg SK, Akturk HK. A New Era in Continuous Glucose Monitoring: Food and Drug Administration Creates a New Category of Factory-Calibrated Nonadjunctive, Interoperable Class II Medical Devices. *Diabetes Technol Ther*. 2018;20(6):391-394.

17. Freckmann G, Nichols JH, Hinzmann R, et al. Standardization process of continuous glucose monitoring: Traceability and performance. *Clin Chim Acta*. 2021;515:5-12.

18. U.S. Food and Drug Administration. Blood Glucose Monitoring Test Systems for Prescription Point-of-Care Use: Guidance for Industry and Food and Drug Administration Staff. Published online 2020.

19. U.S. Food and Drug Administration. Self-Monitoring Blood Glucose Test Systems for Over-the-Counter Use : Guidance for Industry and Food and Drug Administration Staff. Published online 2020.

20. Francq BG, Berger M, Boachie C. To tolerate or to agree: A tutorial on tolerance intervals in method comparison studies with BivRegBLS R Package. *Stat Med*. 2020;39(28):4334-4349.

21. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. CRC Press; 1993.

22. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci*. 1996;11(3):189-228.

23. Shapiro AR. Nonadjunctive Use of Continuous Glucose Monitors for Insulin Dosing: Is It Safe? *J Diabetes Sci Technol*. 2017;11(4):833-838.

24. Krouwer JS. An Analysis of 2019 FDA Adverse Events for Two Insulin Pumps and Two Continuous Glucose Monitors. *J Diabetes Sci Technol*. 2022;16(1):228-232.

Tables

Table 1: FDA requirements defined for point accuracy.[15]

CGM glucose range	*<70 mg/dL (<3.9 mmol/L)*	*70-180 mg/dL (3.9-10.0 mmol/L)*	*>180 mg/dL (>10.0 mmol/L)*	*Total*
*Requirement 1 for agreement rate**	>85%	>70%	>80%	>87%
Limit 1	±15 mg/dL (±0.83 mmol/L)	±15%	±15%	±20%
*Requirement 2 for agreement rate**	>98%	>99%	>99%	-
Limit 2	±40 mg/dL (±2.22 mmol/L)	±40%	±40%	
Requirement 3	No comparator value above 180 mg/dL (10.0 mmol/L	-	No comparator value below 70 mg/dL (3.9 mmol/L)	-

* *Lower bound of one-sided 95% confidence interval*

Table 2: Accuracy parameters of CGM systems A and B.

Comparator glucose range	*<70 mg/dL [b] (<3.9 mmol/L)*		*70-180 mg/dL (3.9-10.0 mmol/L)*		*>180 mg/dL (>10.0 mmol/L)*		*Total*	
CGM system	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Agreement rate [a] ±15 mg/dL (±0.83 mmol/L) *or % [%]*	68.4 (62.8)	71.5 (65.3)	65.7 (62.8)	68.3 (62.5)	74.3 (70.6)	82.3 (79.1)	66.6	71.3
Agreement rate [a] ±20 mg/dL (±1.11 mmol/L) *or % [%]*	80.2	83.6	78.0	78.8	86.5	90.7	78.9 (76.8)	81.3 (77.6)
Agreement rate [a] ±40 mg/dL (±2.22 mmol/L) *or % [%]*	95.8 (93.8)	99.2 (97.8)	96.6 (95.6)	96.0 (94.5)	99.5 (99.1)	99.4 (98.9)	96.6	96.6
Mean absolute relative deviation [%]	23.4	18.9	13.3	13.1	10.6	9.2	13.2	12.3

[a] The values in brackets give the lower one-sided 95% confidence intervals for the agreement rates defined in the FDA requirements.
[b] Absolute deviations were used to calculate the agreement rates

Figures

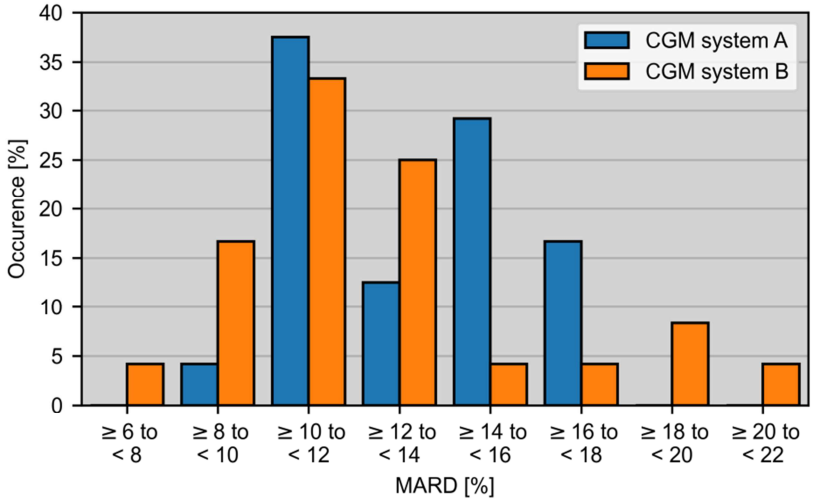

Figure 1: Histogram of MARD results from 24 sensors of CGM systems A (manually calibrated) and B (factory calibrated) each.

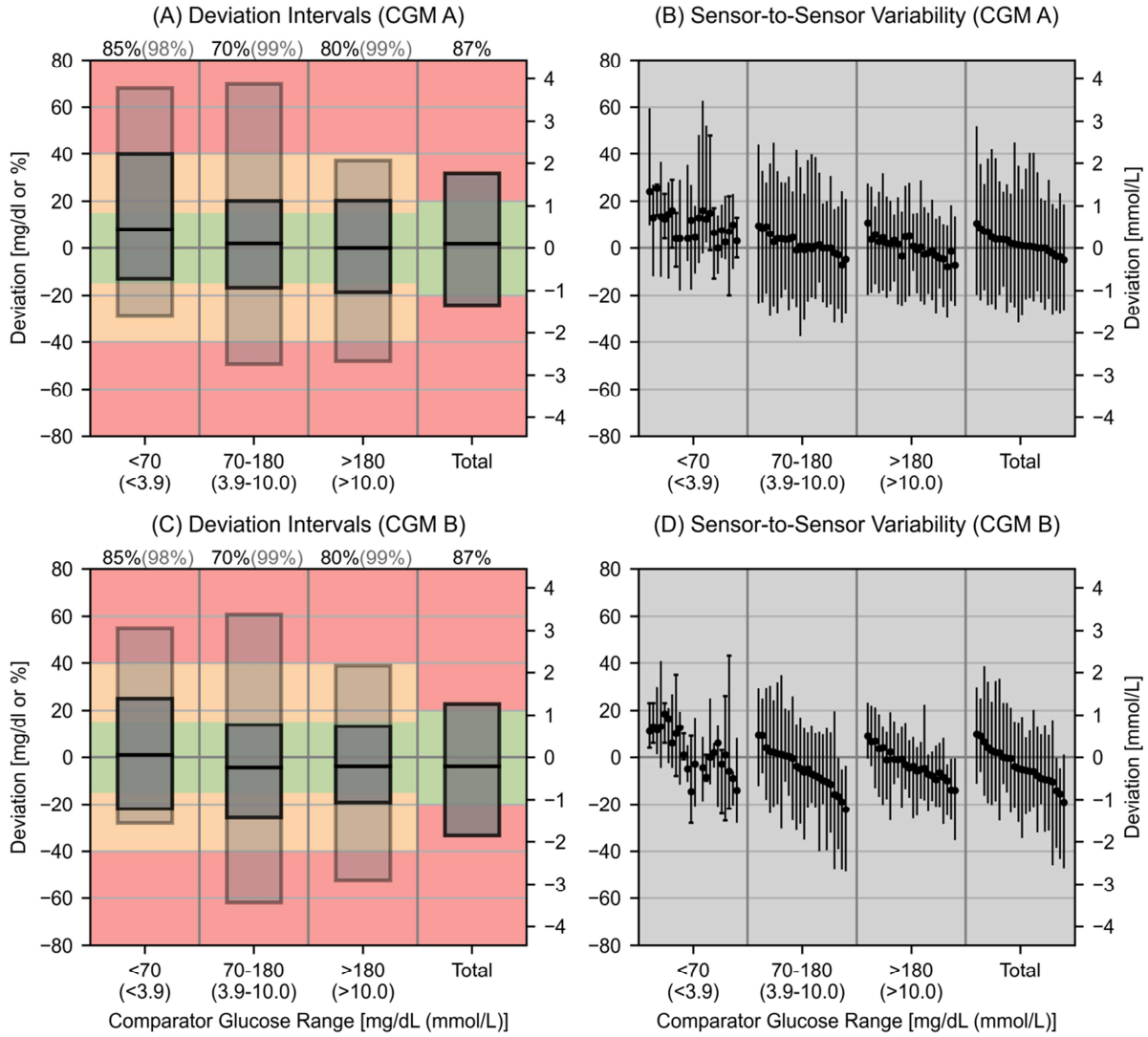

Figure 2: Results of the Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA) for datasets of CGM system A (manually calibrated) (A)-(B) and CGM system B (factory calibrated) (C)-(D). Absolute deviations (in mg/dL on left y-axis and mmol/L on right y-axis) are provided for comparator glucose levels <70 mg/dL (<3.9 mmol/L) and relative deviations for all other levels (in % on left y-axis). (A), (C) Expected ranges of deviations (deviation intervals) in different comparator glucose ranges. The light and dark grey boxes indicate the tolerance intervals (95% confidence) containing the central proportions of deviations printed above the plot. The size of these proportions and the colored background, indicating the extent of deviations, were based on the U.S. Food and Drug administration requirements and limits. (B), (D) Characterization of the sensor-to-sensor variability. Each sensor is described by its median and 90%-range of deviations and sensors are ordered according to the median deviation in the total glucose range. If less than ten data points within a range are available for a sensor, the full range of deviations is displayed and indicated with caps.

Supplementary Material for: "Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA): A Novel Approach for the Statistical Accuracy Assessment of Continuous Glucose Monitoring Systems" by

Eichenlaub, Stephan, Waldenmaier, Pleus, Rothenbühler, Haug, Hinzmann, Thomas, Jendle, Diem and Freckmann

Calculation of tolerance intervals

As mentioned in the main text, the tolerance intervals are calculated using a bias-corrected and accelerated (BCa) bootstrap method accounting for the clustered structure of the data.[21,22] This procedure is carried out as follows:

1. Generate a single bootstrap sample dataset comprised of a randomly sampled selection of sensors (with replacement) and their data containing the same number sensors as the original dataset. As the sampling is done with respect to the sensors and not the individual data points, the clustered structure of the data is accounted for.
2. Calculate the following quantiles of the deviations within the comparator glucose ranges with respect to the corresponding interval sizes.
    a. Comparator glucose <70 mg/dL (<3.9 mmol/L)
        i. Interval 1 (85%) [0.075,0.925]
        ii. Interval 2 (98%) [0.01,0.99]
    b. Comparator glucose 70-180 mg/dL (3.9-10.0 mmol/L)
        i. Interval 1 (70%) [0.15,0.85]
        ii. Interval 2 (99%) [0.005,0.995]
    c. Comparator glucose >180 mg/dL (>10.0 mmol/L)
        i. Interval 1 (80%) [0.1,0.9]
        ii. Interval 2 (99%) [0.005,0.995]
    d. Total comparator glucose
        i. Interval 1 (87%) [0.065,0.935]
3. Repeat steps 1 and 2 until a total of 10,000 bootstrap samples are generated.
4. Calculate the tolerance intervals with 95% confidence by determining the 0.025 quantile of the lower and the 0.975 quantile of the upper interval limits from the

bootstrapped dataset using the BCa method. The BCa method adjusts these quantiles to account for any bias and yield a more reliable estimate of the 95 % confidence intervals.

This procedure was implemented in Python version 3.8 and R version 4.1.2. Using a standard PC (Intel® Core™ i7-770 CPU @ 3.6 GHz, 8GB RAM) the processing time per CGM dataset was approximately 100 seconds in Python and 400 seconds in R.

The reproducibility at a number of bootstrap replications of 10,000 was validated by repeating the entire process 50 times with different seeds for the random generator.

Additional results

Table A1: Deviation intervals for CGM systems A and B

Comparator glucose range	*<70 mg/dL[a] (<3.9 mmol/L)*		*70-180 mg/dL (3.9-10.0 mmol/L)*		*>180 mg/dL (>10.0 mmol/L)*		*Total*	
Interval 1 (2) [%]	*85 (98)*		*70 (99)*		*80 (99)*		*87*	
CGM system	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Upper limit interval 2 [mg/dL (mmol/L) or %]	68.0 (3.8)	55.0 (3.1)	69.8	60.8	37.2	38.9	-	-
Upper limit interval 1 [mg/dL (mmol/L) or %]	39.9 (2.2)	25.0 (1.4)	20.2	13.8	20.3	13.2	31.7	22.8
Median [mg/dL (mmol/L) or %]	8.0 (0.4)	1.0 (0.1)	1.9	-4.3	0.0	-3.8	1.8	-3.9
Lower limit Interval 1 [mg/dL (mmol/L) or %]	-13.0 (-0.7)	-22.0 (-1.2)	-16.8	-25.8	-18.7	-19.1	-24.2	-33.3
Lower limit Interval 2 [mg/dL (mmol/L) or %]	-29.0 (-1.6)	-28.0 (-1.6)	-49.3	-61.9	-48.0	-52.2	-	-

[a] Absolute deviations