

**This manuscript has been published in:**

Journal of Diabetes Science and Technology 1–9

DOI: 10.1177/19322968221134639

## **Title**

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

## **Authors**

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

- 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 (WG-CGM)

## **Abbreviations**

CGM                                      Continuous Glucose Monitoring

ADA	American Diabetes Association
MARD	Mean Absolute Relative Difference
AR	Agreement Rate
BGMS	Blood Glucose Monitoring System
PARD	Paired Absolute Relative Difference
CLSI	Clinical and Laboratory Standards Institute
FDA	United States Food and Drug Administration
CG-DIVA	Continuous Glucose Deviation Interval and Variability Analysis
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine

### **Key words**

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

### **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

### **Funding Source**

This study was supported by Diabetes Center Berne, Switzerland.

### **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.

## **Acknowledgements**

The authors would like to thank the members of the working group on continuous glucose monitoring established 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.

## **Number of words, figures and tables**

Abstract: 248, Main text: 3665

Figures: 2, Tables: 3

## **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 which is based on the U.S. Food and Drug Administration requirements for “integrated” CGM systems.

### *Methods*

The statistical concept of tolerance intervals and data from two approved CGM systems was used to illustrate the CG-DIVA.

### *Results*

The CG-DIVA characterizes the expected range of deviations of the CGM system from a comparison method in different glucose concentration ranges and 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/lfDTUlm/CGM\\_Performance\\_Assessment](https://github.com/lfDTUlm/CGM_Performance_Assessment)).

### *Conclusions*

We argue that the CG-DIVA can simplify 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

Longstanding research in continuous glucose monitoring (CGM) has highlighted 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 assumed to be of higher measurement quality at a single point in time, as opposed to trend accuracy focusing on rate of change agreement.

Many different approaches to characterize CGM point accuracy have been utilized over the years. Despite ongoing criticism,<sup>1–4</sup> 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 certain limits, e.g.  $\pm 20\%$ , of their paired comparator measurements. Additionally, the calculation of correlation coefficients in the context of linear regression analysis is common. Furthermore, the error grid analysis, originally developed for blood glucose monitoring systems (BGMS) but also applied to CGM systems, is very common. It provides another parameter as well as a graphical visualization.<sup>5–7</sup> A dedicated error grid analysis for CGM systems that includes the rate of change in glucose levels has been proposed in 2004.<sup>8</sup> However, as its calculation and interpretation are considered to be very complex, it is rarely used today.<sup>9,10</sup> To determine CGM precision, the mean deviation between paired measurements of two sensors in the same subject (PARD) is commonly used. However, a more useful characterization of the variability of accuracy between sensors can be provided with a histogram of sensor-specific MARD values.<sup>11–13</sup>

From this brief overview of parameters, it should become clear that the discussion of CGM accuracy is complicated by the abundance of existing approaches.<sup>10</sup> In this context, the Clinical and Laboratory Standards Institute (CLSI) has published the POCT05 guideline on “performance metrics for continuous interstitial glucose monitoring”,<sup>14</sup> which seems to be at least partially implemented in user manuals of current CGM systems.<sup>15,16</sup> Here, extensive tables with AR results for different limits (e.g.  $\pm 15\%$ ,  $\pm 20\%$ ,  $\pm 30\%$  and  $\pm 40\%$ ) and MARDs in various glucose ranges are provided and accompanied by concurrence tables between CGM

and comparator measurements in an even larger number of glucose ranges. This, however, fails to alleviate the previously mentioned issue of abundance and it is questionable whether this method of accuracy characterization is suitable for the readers of user manuals, i.e., clinicians and patients.

In 2018, the United States Food and Drug Administration (FDA) released special control regulations, which, for the first time, included the definition of minimum acceptance requirements for “integrated” CGM systems (Table 1).<sup>17,18</sup> 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. However, ARs alone fail to provide separate information on bias and imprecision and a distinct analysis of sensor-to-sensor variability is not demanded at all.

With the aim of further standardizing and simplifying the characterization of CGM accuracy while also increasing its accessibility, this article introduces the Continuous Glucose Deviation Interval and Variability Analysis (CG-DIVA). 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 mainly appeal to manufacturers, researchers and regulators but clinical practitioners should also be able to interpret it.

The present article supports the activities of a working group on CGM (WG-CGM) established by the Scientific Division of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC).<sup>19</sup>

## Methods

### *Statistical methodology*

The CG-DIVA focuses on the analysis of deviations between paired CGM and comparator measurements, which can provide independent information on precision and bias. A suitable approach for describing the distribution of these deviations would be to use the well-known boxplot where the median (representing bias) and the central 50% interval between the 25<sup>th</sup> and 75<sup>th</sup> percentiles (representing precision) of the deviations are calculated. To consolidate this approach with the FDA requirements for point accuracy (Table 1), the statistical concept of tolerance intervals was utilized,<sup>20</sup> as it has two main advantages over the classical boxplot. Firstly, it allows the calculation of an interval by which any percentage of deviations is covered, not just 50% in case of the boxplot. We can therefore use of the minimum requirements for the ARs, e.g. 87%, to define the coverage of the intervals. Secondly, tolerance intervals incorporate the statistical uncertainty that comes with any data collection process, thus implementing the FDA requirement of using confidence intervals on their ARs. In particular, the tolerance intervals make a statement about the range of deviations generally expected from the CGM system (with a confidence of 95%), not just the particular range found in the dataset of one clinical study.

For the calculation of the tolerance intervals a bias-corrected and accelerated bootstrap method<sup>21,22</sup> 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 novel approach was compared with conventional accuracy parameters by calculating ARs and their lower 95% confidence limits (using the same bootstrapping procedure as described above) as well as aggregate MARDs (averaged over all CGM-comparator pairs).

### *Data description*

A detailed description of the study used to exemplify the CG-DIVA has been published previously.<sup>23,24</sup> In short, data was collected from 24 subjects using two approved CGM devices. CGM system A required manual calibrations, whereas system B was factory-calibrated. The systems are pseudonymized because this article focuses on the CG-DIVA as a methodology and not the performance of the CGM systems, which has been discussed previously.<sup>23,24</sup> Comparator measurements were obtained from capillary samples using a commercially available blood glucose monitoring system (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 in the ranges <70, 70-180 and >180 mg/dL (<3.9, 3.9-10.0 and >10.0 mmol/L), respectively.



## Results

### *Conventional accuracy characterization*

The results of a conventional accuracy characterization including ARs and MARDs are provided in Table 2. Furthermore, a histogram of individual sensor MARDs 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 for point accuracy in Table 1. However, the CG-DIVA divides the glucose range according to the comparator, not CGM measurements, 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 two parts. Its results are displayed in two separate plots (Figure 2) and in a table (Table 3) described below.

### *Deviation intervals*

The first part of the CG-DIVA uses the previously described concept of tolerance intervals to infer so-called deviation intervals, i.e., the range of deviations from a comparator that can be expected from the CGM system in general. They are depicted by the boxes in Figure 2 (A) and (C). The median deviations are shown by the black dashes, thus a dash above zero (positive median deviation) would mean a systematic overestimation of glucose levels and vice versa. The coverage of the deviation intervals, i.e., the percentage of deviations falling within the intervals were set according to the AR requirements defined by the FDA (Table 1) and printed

above the plot. The dark grey box represents FDA requirement 1 and the light grey box represents requirement 2. The colored background indicates the extent of the deviations and was based on limit 1 and 2 of the FDA requirements. The corresponding values for the limits of the deviation intervals as well as the range they cover are provided in Table 3. For example, the CG-DIVA results for CGM system A can be interpreted as follows: When glucose levels are  $>180$  mg/dL ( $>10.0$  mmol/L) we can expect that 80% of deviations lie between  $+20.3\%$  and  $-18.7\%$  (dark grey box). As this interval is centered around the median, it directly follows that 10% of deviations are expected to be greater than  $+20.3\%$  and the remaining 10% of deviations greater than  $-18.7\%$ . Analogously to that, the second interval (light grey box) gives the information that 99% of deviations are expected to lie between  $+37.2\%$  and  $-48.0\%$  and therefore, 0.5% or 1 in 200 deviations are expected to fall above and below this interval, respectively.

Additionally to this information on the extent of deviations, the CG-DIVA results can also be interpreted in relation to the corresponding ARs and their lower confidence bounds in Table 2. 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. the upper and lower 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, like 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.

#### *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. Specifically, a positive median deviation means that the sensor systematically overestimates the glucose level and vice versa. The sensors are ordered by their median deviations across the total glucose range. If less than ten data points are available for a sensor, which only occurs in the glucose range  $<70$  mg/dL ( $<3.9$  mmol/L) in the examples, 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. The corresponding numerical results of between and within sensor variability are summarized in Table 3 with the minimum/maximum and range of median sensor deviations as well as the median of 90% ranges, respectively. In contrast to the first part of the CG-DIVA, it should be emphasized that this is a purely descriptive analysis of the collected data not accounting for the statistical uncertainty.

## Discussion

The CG-DIVA is based on the FDA requirements for integrated CGM systems,<sup>17,18</sup> and 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. The properties of the dataset used in this article can be considered as a minimum with regards to number of sensors and number of data points per comparator glucose range. The example datasets presented in this article were chosen to explain and illustrate the CG-DIVA and its interpretation rather than to discuss accuracy of the examined CGM systems.

Diverging from the FDA requirements, the CG-DIVA uses the comparator instead of CGM measurements to stratify glucose levels into different ranges. This choice was made as the stratification with respect to the examined measurement system 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.<sup>25,26</sup> Furthermore, it means that the method for stratifying the data 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 hinders the comparison of different CGM systems amongst each other.

In comparison to the conventional ARs and MARDs, the deviation intervals in the CG-DIVA can provide a more comprehensive 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) and Table 3, 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,<sup>1-4</sup> 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,<sup>27,28</sup> 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 ( $\pm 15$  and  $\pm 40$  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 the 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 and not demanded by the FDA. Examining the conventional approach of displaying a histogram of MARDs in Figure 1, it can be suspected that CGM system B has a higher variability between sensors; however, the impact of this variability on the measurement results is not clear. In contrast, Figure 2 (B) and (D) and Table 3 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 within sensor variability (90% range indicated by the antennae in the plot range and summarized in the last row of Table 3) 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.<sup>29</sup> While these ranges are well-established and can be used to characterize the status of glycemic control, they are not ideal 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 glucose levels are decreasing. 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).

An additional improvement could be made by adapting the coverage percentages of the deviation intervals, i.e. the share of deviations they display. In its current form, the coverage percentages are different for each comparator glucose range, thus making it difficult to compare the accuracy, in particular imprecision, across ranges. These percentages were chosen according to FDA requirements 1 and 2 for the ARs, however the clinical rationale for these requirements is unclear. 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 coverage of deviation intervals 1 and 2 to 80% and 98%, respectively. This, however, would also require an adaptation of the acceptance limits, e.g.  $\pm 15\%$  or  $\pm 40\%$ , depicted with the colored background of the deviation interval plot.

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 trend accuracy requires frequent comparator sampling over longer periods of time. Furthermore, the study design, in particular the fraction of data points obtained during rapid changes in glucose concentration, 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/lfdTUlm/CGM\\_Performance\\_Assessment](https://github.com/lfdTUlm/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, especially its graphical display, represents an attractive alternative to the conventional approach of tabulating point accuracy parameters such as MARD and AR in extensive tables, thus 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 coverage 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 a 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 (WG-CGM) of the IFCC.



## References

1. 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.
2. 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.
3. 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.
4. 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.
5. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. *Diabetes Care*. 1987;10(5):622-628.
6. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*. 2000;23(8):1143-1148.
7. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014;8(4):658-672.
8. 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.
9. Wentholt IM, Hoekstra JB, DeVries JH. A Critical Appraisal of the Continuous Glucose-Error Grid Analysis. *Diabetes Care*. 2006;29(8):1805-1811.
10. Bailey TS. Clinical Implications of Accuracy Measurements of Continuous Glucose Sensors. *Diabetes Technol Ther*. 2017;19(S2):S-51.
11. Bailey T, Bode BW, Christiansen MP, Klaff LJ, Alva S. The Performance and Usability of a Factory-Calibrated Flash Glucose Monitoring System. *Diabetes Technol Ther*. 2015;17(11):787-794.
12. Garg SK, Kipnes M, Castorino K, et al. Accuracy and Safety of Dexcom G7 Continuous Glucose Monitoring in Adults with Diabetes. *Diabetes Technol Ther*. 2022;24(6):373-380.
13. Shah VN, Laffel LM, Wadwa RP, Garg SK. Performance of a Factory-Calibrated Real-Time Continuous Glucose Monitoring System Utilizing an Automated Sensor Applicator. *Diabetes Technol Ther*. 2018;20(6):428-433.
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. Dexcom. Dexcom G6 Continuous Glucose Monitoring System User Guide. Published online 2022. <https://s3-us-west-2.amazonaws.com/dexcompdf/G6-CGM-Users-Guide.pdf>
16. Abbott. Freestyle Libre 2 Flash Glucose Monitoring System User Guide. Published online 2020.

[https://freestyleserver.com/distribution/fxaa20.aspx?product=ifu\\_art44417\\_001&version=latest&os=all&region=us&language=xx\\_yy](https://freestyleserver.com/distribution/fxaa20.aspx?product=ifu_art44417_001&version=latest&os=all&region=us&language=xx_yy)

17. 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](https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170088.pdf)
18. 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.
19. Freckmann G, Nichols JH, Hinzmann R, et al. Standardization process of continuous glucose monitoring: Traceability and performance. *Clin Chim Acta*. 2021;515:5-12.
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. Pleus S, Stuhr A, Link M, Haug C, Freckmann G. Variation of Mean Absolute Relative Differences of Continuous Glucose Monitoring Systems Throughout the Day. *J Diabetes Sci Technol*. 2022;16(3):649-658.
24. Freckmann G, Pleus S, Schauer S, et al. Choice of Continuous Glucose Monitoring Systems May Affect Metrics: Clinically Relevant Differences in Times in Ranges. *Exp Clin Endocrinol Diabetes Off J Ger Soc Endocrinol Ger Diabetes Assoc*. 2022;130(5):343-350.
25. 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.
26. 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.
27. Shapiro AR. Nonadjunctive Use of Continuous Glucose Monitors for Insulin Dosing: Is It Safe? *J Diabetes Sci Technol*. 2017;11(4):833-838.
28. 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.
29. 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.

## Tables

**Table 1:** FDA requirements for point accuracy.<sup>17</sup>

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 <sup>b</sup> (<3.9 mmol/L)		70-180 mg/dL (3.9-10.0 mmol/L)		>180 mg/dL (>10.0 mmol/L)		Total	
	A	B	A	B	A	B	A	B
CGM system								
Agreement rate [%] <sup>a</sup> ±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 [%] <sup>a</sup> ±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 [%] <sup>a</sup> ±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
Aggregate mean absolute relative difference [%]	23.4	18.9	13.3	13.1	10.6	9.2	13.2	12.3

<sup>a</sup> The values in square brackets give the lower one-sided 95% confidence intervals for the agreement rates defined in the FDA requirements.

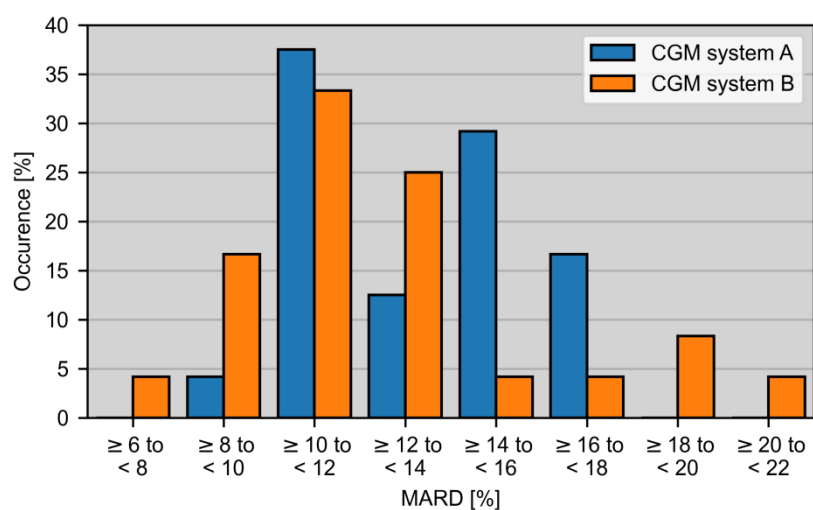
<sup>b</sup> Absolute deviations were used to calculate the agreement rates

**Table 3:** Numerical results of the CG-DIVA. Deviation intervals are quantified by their limits and their range (distance between lower and upper limit). The quantifiers of between sensor variability give the minimum and maximum of individual sensor median deviations as well as their range. The quantifier of within sensor variability gives the median of individual sensor 90% ranges (distance between lower and upper limit).

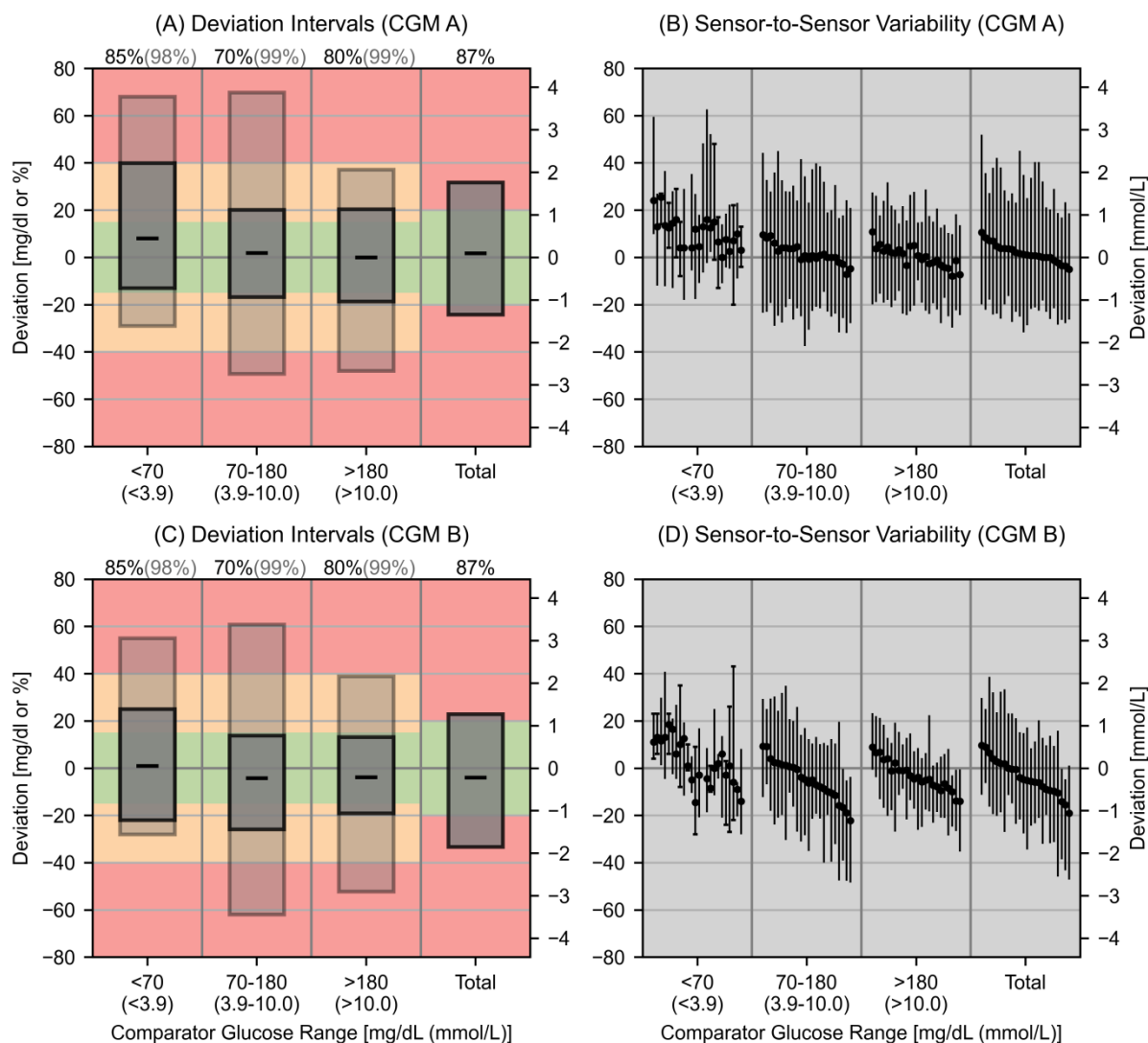
Comparator glucose range	<70 mg/dL <sup>a</sup> (<3.9 mmol/L)		70-180 mg/dL (3.9-10.0 mmol/L)		>180 mg/dL (>10.0 mmol/L)		Total	
CGM system	A	B	A	B	A	B	A	B
Coverage of deviation Intervals 1 and 2 [%]	85 and 98		70 and 99		80 and 99		87	
Median [mg/dL (mmol/L) or %]	8.0 (0.4)	1.0 (0.1)	1.9	-4.3	0.0	-3.8	1.8	-3.9
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
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
Range interval 1 [mg/dL (mmol/L) or %]	52.9 (2.9)	47.0 (2.6)	37.0	39.6	39.0	32.3	55.9	56.2
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	-	-
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	-	-
Range interval 2 [mg/dL (mmol/L) or %]	97.0 (5.4)	83.0 (4.6)	119.1	122.7	85.3	91.1	-	-
Between sensor variability (min – max) [mg/dL (mmol/L) or %]	[0.0 - 25.5] ([0.0 - 1.4])	[-14.5 - 18.5] ([-0.8 - 1.0])	[-7.2 - 9.6]	[-22.2 - 9.3]	[-8.0 - 10.8]	[-14.0 - 8.9]	[-5.1 - 10.5]	[-19.0 - 9.7]
Between sensor variability (range) [mg/dL (mmol/L) or %]	25.5 (1.4)	33.0 (1.8)	16.8	31.5	18.8	22.9	15.6	28.7
Within sensor variability [mg/dL (mmol/L) or %]	41.0 (2.3)	30.9 (1.7)	55.1	44.9	40.5	27.7	52.5	41.9

<sup>a</sup> Absolute values in mg/dL (mmol/L)

## Figures



**Figure 1:** Histogram of mean absolute relative deviations (MARD) calculated over each of the 24 sensors of CGM systems A (manually calibrated) and B (factory calibrated) each. Aggregate MARD results (average of all CGM-comparator pairs) are 13.2% and 12.3% for CGM systems A and B respectively.



**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) Deviation intervals, i.e., the expected range of deviations, in different comparator glucose ranges. The median deviations are shown by the black dashes. The coverage of the deviation intervals, i.e., the percentage of deviations falling within the intervals were set according to the requirements defined by the U.S. Food and Drug administration (FDA) and printed above the plot. The dark grey box represents FDA requirement 1 and the light grey box represents requirement 2. The colored background indicates the extent of the deviations and was based on limit 1 and 2 of the FDA requirements. (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.