Review Article

Recommendations on the collection of comparator measurement data in the performance evaluation of continuous glucose monitoring systems

Guido Freckmann[1,2,] Stefan Pleus[1,2*], Manuel Eichenlaub[2], Elisabet Eriksson Boija[1,3], Marion Fokkert[1,4], Rolf Hinzmann[1,5], Johan Jendle[1,6], David C. Klonoff[1,7], Konstantinos Makris[1,8], James H. Nichols[1,9], John Pemberton[1,10], Elizabeth Selvin[1,11], Nam K. Tran[1,12], Lilian Witthauer[1,13,14], Robbert J. Slingerland[1,4]

on behalf of the Working Group on Continuous Glucose Monitoring of the IFCC Scientific Division

[1]*IFCC Scientific Division - Working Group on Continuous Glucose Monitoring*

[2]*Institut für Diabetes-Technologie, Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm, Ulm, Germany*

[3]*Equalis AB, Uppsala, Sweden*

[4]*Department of Clinical Chemistry, Isala Clinics, Zwolle, the Netherlands*

[5]*Lampertheim, Germany*

[6]*School of Medical Sciences, Faculty of Medicine, and Health, Örebro University, Örebro, Sweden*

[7]*Diabetes Research Institute of Mills-Peninsula Medical Center, San Mateo, CA, USA*

[8]*Clinical Biochemistry Department, KAT General Hospital, Athens, Greece*

[9]*Vanderbilt University Medical Center, Nashville, TN, USA*

[10]*Birmingham Women's and Children's Foundation Trust, Birmingham, United Kingdom*

[11]*Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

[12]*Department of Pathology and Laboratory Medicine, University of California Davis Health, Sacramento, CA, USA*

[13]*Department of Diabetes, Endocrinology, Nutritional Medicine, and Metabolism, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland*

[14]*Diabetes Center Berne, Bern, Switzerland*

Article Info

*Corresponding Author:
Stefan Pleus, PhD
Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm
Lise-Meitner-Strasse 8/2, D-89081 Ulm, Germany
E-mail: stefan.pleus@idt-ulm.de
Phone +49-731-50-990-0

Abstract

While current systems for continuous glucose monitoring (CGM) are safe and effective, there is a high degree of variability between readings within and across CGM systems. In current CGM performance studies, device readings are compared to glucose concentrations obtained with a comparator ("reference") measurement procedure (usually capillary or venous glucose). However, glucose concentrations from capillary and venous samples can systematically differ, often by as much as 5 to 10%. Different comparator methods have shown biases of up to 8%, and comparator devices of the same brand can systematically differ by more than 5%. To address these issues, the Working Group on CGM of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC WG-CGM) recommends standardizing study procedures and the comparator measurement process in CGM performance studies. The majority of IFCC WG-CGM members recommend the use of capillary samples as reference, mainly because CGM readings will then be aligned better with results from self-monitoring of blood glucose (SMBG). Even with factory-calibrated CGM systems, manufacturers require CGM users to perform SMBG in some situations, e.g., manual calibration, confirmation of extreme readings, discordance between CGM readings and symptoms of hyper- or hypoglycemia, or intermittent signal loss.

Comparator devices should meet defined analytical performance specifications for bias and imprecision. Comparator bias can be reduced by retrospective correction of comparator values based on measurements with a method or materials of higher metrological order. Once manufacturers align CGM readings of their systems with comparator results using standardized procedures, variability across CGM systems will be reduced.

Introduction

Continuous glucose monitoring (CGM) is an essential tool for many people with diabetes. Studies have shown the clinical benefit of CGM systems in persons with type 1 diabetes, independent of being used as a stand-alone device or integrated with sensor-augmented pumps or systems for automated insulin delivery, and there is growing evidence for potential benefit in persons with type 2 diabetes [1,2].

CGM systems can display markedly different readings when used in parallel, even if they show good performance results in their pivotal study [3-10]. As a result, variability between CGM systems can be substantial enough that two systems may give sufficiently different results to affect clinical decisions.

The Working Group on CGM of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC WG-CGM) aims to standardize procedures for assessing the performance of CGM systems. A standardized study design with harmonized measurement procedures will allow better direct comparisons of CGM performance, and readings from different CGM systems will be better aligned. The IFCC WG-CGM acknowledges that procedures may have to be modified for specific populations (e.g., people <18 years of age) or for specific indications (e.g., intensive care units).

Specific characteristics for comparator data have been recommended before:[11] Comparator data should be obtained at 15-minute intervals for 6 to 8 hours during the beginning, the middle, and the end of sensor lifetime. Furthermore, the comparator data distribution should cover the clinically relevant concentration ranges and rates of change, for example, by inducing glucose changes after meal challenges.

To ensure comparability, the procedure for comparator ("reference") measurements must be standardized. A recent review indicates that this has not been the case in the past [12]. The complete description of a comparator measurement procedure comprises the type of comparator device (e.g., laboratory analyzer, blood gas analyzer, handheld analyzer), the brand of the device, the device's analytical methodology (e.g., glucose oxidase, hexokinase), its traceability to a higher order reference method (e.g., isotope dilution mass spectrometry [IDMS]), the specific consumable batches/reagent lots used, and the type of sample (sample origin and pre-analytical handling). The choice of sample type and comparator method can affect each other, as not all methods are intended to be used with all sample types.

This article reports recommendations by the IFCC WG-CGM for standardized study procedures. With these recommendations, the IFCC WG-CGM intends to initiate scientific discourse regarding design of and procedures in CGM performance studies. Ultimately, the IFCC WG-CGM aims to support the establishment of an international standard regarding the performance evaluation of CGM systems.

Sample origin and handling

Background information on sample origin

In CGM performance studies, three blood sample types are commonly used for comparator measurements: capillary, venous, and "arterialized-venous" [12]. Glucose concentrations are physiologically different between these venous and capillary samples, with past studies showing differences between capillary and venous glucose concentrations of, on average, 5% to 10% that can vary depending on the glucose rate of change. While arterialized-venous samples are also drawn from the vein, they undergo different pre-analytical handling by warming the arm or hand where the blood is drawn. In a recent study [13], the difference between capillary and venous glucose concentrations was +5.9%, which could be reduced to +4.2% by applying heat to the arm. As a result, the observed CGM performance was different when using venous, arterialized-venous, or capillary glucose concentrations [13]. The bias of the investigated CGM system was approximately +2% when comparing against venous comparator samples, +0% against arterialized-venous comparator samples, and -4% against capillary comparator samples. This resulted in mean absolute relative difference (MARD) results of approximately 7.5% (venous comparator samples), 7% (arterialized-venous comparator samples), and 8% (capillary comparator samples). It is, therefore, imperative to standardize the origin of samples for comparator measurements in CGM performance studies. Table 1 summarizes advantages and disadvantages of the different sample origins. It should be noted that the selection of sample origin and the selection of comparator measurement method affect each other.

Table 1: Summary of sample origins with associated advantages and disadvantages.

	Advantages	Disadvantages
Capillary	• Relied on by users to verify CGM accuracy • Used when CGM data is unavailable or implausible • Used for manual CGM calibration • Sampling has low risk of adverse event • Immediate display of results for glucose manipulation procedure if BGMS is used • Considered acceptable in all age groups, including children ≤6 years of age • Does not require sampling by trained health care professional • Alignment of in-clinic and free-living data[3]	• Repeated fingersticks can be painful and burdensome • Sample volume may be too small for some glucose analyzers accepted as comparator by the FDA[1] • Compliance with analytical performance specifications and retrospective correction may have to be performed in two steps (vs. lab analyzer and lab analyzer vs. higher-order material) if higher-order material cannot be used with the comparator device[2]
Venous	• Sample volume sufficient for glucose analyzer accepted as comparator by the FDA • Usually, only a single stick is required to place a peripheral or central intravenous line. • Clinical diagnostics are typically based on venous glucose concentrations due to ease of access. • In a situation requiring an urgent result, venous blood from an indwelling intravenous line is the most reliably obtained type of specimen.	• Sampling and sample handling require trained healthcare professionals. • Increased risk of adverse events due to invasiveness • Total blood loss might be an issued • Sampling technique and sample handling can impact measurement results (inadvertent hemolysis, glycolysis) • Immediate display of results for monitoring of glucose manipulation procedure is challenginge • Not feasible in a free-living setting • Not feasible for minors ≤6 years of age
Arterialized-venous	• See venous advantages • Better alignment with capillary glucose levels than venous glucose levels, i.e., results of in-clinic and free-living procedures are more comparable[3]	• See venous disadvantages • **Heating procedure currently is not standardized[7]** • Need for additional resources to ensure sustained levels of heat application[8] • Immobilization of participants • Additional burden on participants due to prolonged heat application

CGM: continuous glucose monitoring system, FDA: Food and Drug Administration, BGMS: blood glucose monitoring system

Information is based on the most common measurement schedule,[11] which is also recommended by the IFCC WG-CGM, which requires blood sampling in 15-minute intervals over a period of 6 to 8 hours, i.e., 25 to 33 consecutive blood draws.

[1]Many laboratory-grade analyzers need at least 25 µL of sample volume. While whole blood samples of this volume might be obtained with the proposed sampling frequency and duration of sampling periods, plasma samples would need additional whole blood to be collected, which cannot be achieved reliably.

[2]Analytical performance specifications and retrospective correction are addressed in section "Analytical performance of comparator methods". It may be more difficult for POC-BGMS or SMBG systems to fulfill C-APS than for laboratory analyzers.

[3]Manufacturers likely base their algorithms on both in-clinic and free-living or real-world data. Alignment leads to better comparability between these data.

[4]The Clinical and Laboratory Standards Institute's (CLSI) guideline POCT5[14] recommends collecting no more than 5 mL/kg body weight of blood within any 24-hour period or 7 mL/kg in any eight-week period. These limits are lower in protected populations, like pregnant or pediatric participants. It should be noted that the actual blood volume drawn depends on the sampling technique, e.g., loss due to line flushing, and sample container size.

[5]Venous samples have to be collected and centrifuged before measurements can be performed. There will be a delay before display of results.

[6]The arterialization effect likely depends on aspects of the specific heating procedure, e.g., arm or hand to be heated, type of heating pad or chamber, temperature setting. It would, therefore, be necessary to describe the heating procedure in much more detail and to investigate how the arterialization effect depends on the heating procedure.

[7]Additional resources may include temperature monitoring to minimize the risk of excessive temperatures and restarting the heating pad, which are typically automatically switched off after some time for safety reasons. In addition, and this overlaps with the immobilization of participants, participants need help removing and reapplying the heating pad around restroom breaks.

While glucose concentrations also depend on the sample matrix (i.e., whole blood, plasma, serum), whole-blood glucose concentrations can be converted to plasma-equivalent values so that any remaining differences are negligible [15]. However, it must be noted that a conversion factor is only an approximation and the individual association between whole-blood and plasma glucose may be influenced by other factors, like hematocrit and illness [16,17]. Measurement of glucose concentrations in serum is inadequate in the context of comparator data in CGM performance studies due to glycolysis.

Recommendation for sample origin

The majority of WG-CGM members recommend capillary samples to be the preferred sample origin in CGM performance studies. Since capillary self-monitoring is supplemented by adjunctive CGM systems and the fallback/verification option for nonadjunctive CGM systems, using capillary comparator samples would optimize the clinical alignment of CGM readings and self-monitoring of blood glucose (SMBG) concentrations. The IFCC WG-CGM viewed this alignment as more important for end-users than the alignment of CGM readings to venous glucose concentrations, which are used in the diagnosis of diabetes and for diabetes management in healthcare settings. However, the IFCC WG-CGM acknowledges that during the development of a future international standard, which is an aim of the IFCC WG-CGM, the sample type must be discussed with different stakeholders like diabetologists, regulatory authorities, and CGM manufacturers.

Analytical performance of comparator methods

Background information

The consistent analytical performance of comparator methods is essential in CGM performance studies not only to optimize comparability of results between studies, but also because the comparator is assumed to be free from error. For an accuracy metric like measurement bias between a CGM system and a comparator to be meaningful, there is an assumption that the comparator indicates the "true" glucose value. In reality, however, any measurements, including comparator measurements, are affected by different sources of error. Untangling comparator and CGM error is not feasible. Any inaccuracies of the comparator will affect the performance of a CGM system as observed in a study, and so will any differences in comparator method across different studies.

Of course, comparators must be used in compliance with device labeling and applicable legal requirements. Still, comparators can exhibit measurement bias of up to 8% depending on the measurement method [18,19]. In addition, more than 5% bias between individual devices of the same brand can exist [20,21]. Analytical performance of comparators can be optimized by requiring that comparator methods have established metrological traceability with defined analytical performance specifications. These specifications are called "analytical performance specifications for comparators" (C-APS) here to avoid confusion with minimum requirements for CGM performance. C-APS comprise bias and imprecision. C-APS can be defined based on biological variation [22,23]. It may be more difficult for POC-BGMS or SMBG systems to fulfill C-APS than for laboratory analyzers, although it has been shown that they can satisfy the C-APS [20,24]. Furthermore, they may not be recognized by all regulatory bodies. The FDA, for example, stated in the past that they only recognize "laboratory-based methods" [25]. In the European Union, however, a CGM system gained CE marking based on capillary comparator data obtained with an SMBG system [26]. As long as the comparator is used within the intended use and performs within C-APS, the specific brand or type of device (laboratory analyzer, POC-BGMS or SMBG system) is irrelevant, especially if bias is minimized through retrospective correction. It should be noted that compliance with C-APS does not reflect potentially interfering substances. The effect of interfering substances on laboratory analyzers, POC-BGMS, and SMBG systems typically increases in that order, although it depends on the specific brand. This could limit the choice of the comparator in CGM performance studies.

Assessment of measurement bias

Bias is defined as the estimate of a systematic difference between methods. In the context of comparators, bias should be assessed using a method or materials of higher metrological order. The higher metrological order associated with these methods and materials implies that corresponding measurement results and assigned target values, respectively, are closer to the "true" glucose concentration than the measurement results obtained with the comparator. An IDMS-based method could constitute a method of higher metrological order if a laboratory analyzer is selected as comparator, and bias could

be verified with a split-sample approach, where comparator samples are split into two aliquots, one for the comparator and one for the higher-order method. If professional-use point-of-care (POC) blood glucose monitoring systems (BGMS) or "over the counter" systems for SMBG are used as comparator, then an additional step by way of, for example, a laboratory analyzer may be required if materials of higher order are used to assess bias, but do not fall within the intended use of the POC-BGMS or SMBG system. In case higher-order materials are used, they are ideally independent from the manufacturer of the selected device. For example, the serum-based Standard Reference Material 965 (National Institute of Standards and Technology, Gai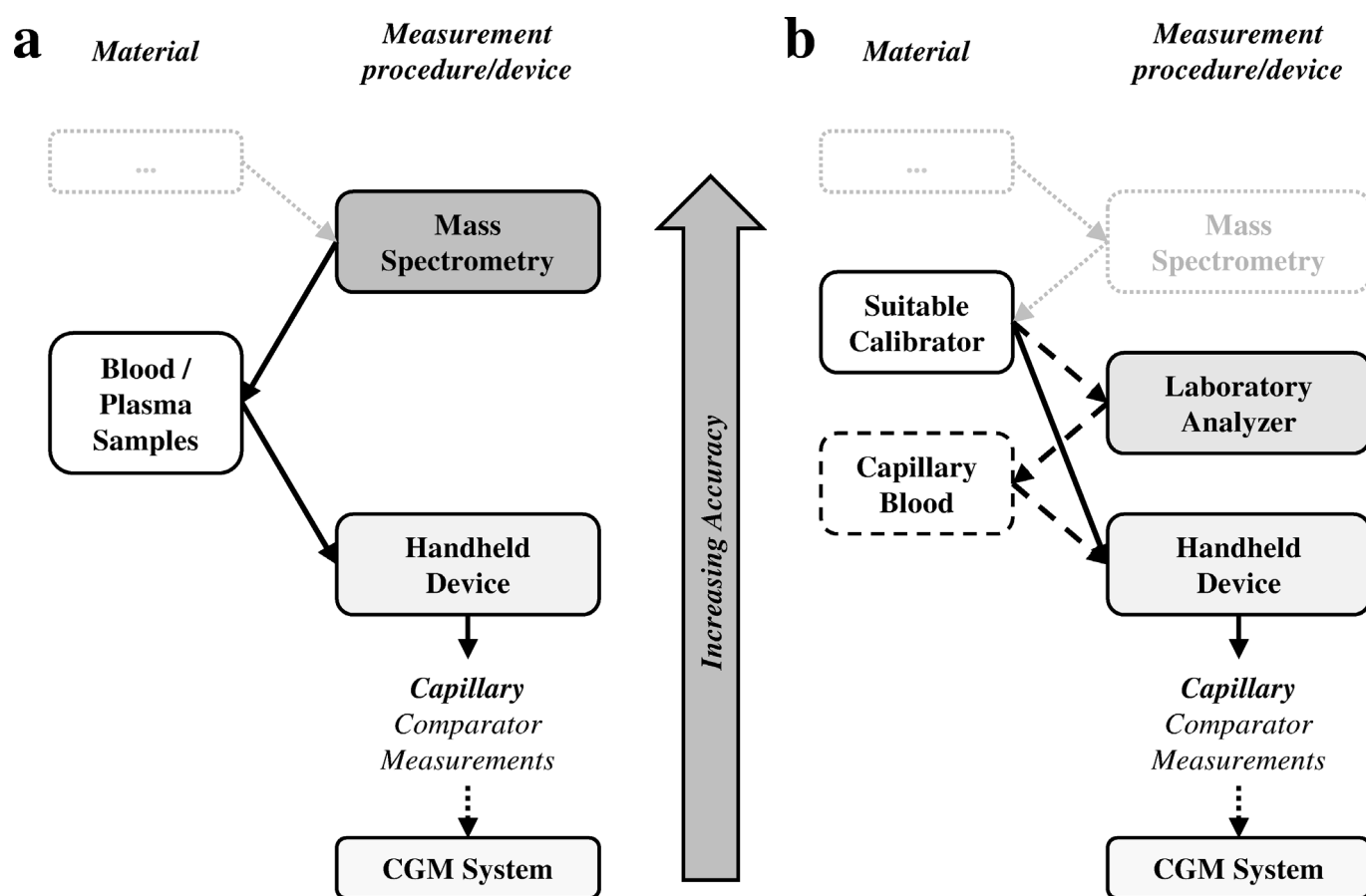thersburg, MD) could serve as a material of higher metrological order if the samples are commutable when measured with the specific comparator. The steps associated with bias verification are shown in Figure 1.

Commutability of the samples used for bias verification is important. In lay-person's terms, commutability means that the higher-order materials behave like participant samples (whole blood or plasma) in both the method used to assign the target concentrations and the comparator method used in the performance study. Therefore, participant samples in a split-sample approach are commutable by definition, whereas commutability of other samples may have to be verified in advance.

Figure 1: Example of a traceability chain when estimating bias.
Panel a) Estimation based on methods of higher metrological order.
Panel b) Estimation based on materials of higher metrological order ("suitable calibrator").

If participant blood/plasma samples are used to verify the comparator method bias, potential commutability issues can be avoided. When using capillary comparator measurements, an additional step (indicated by dashed lines and boxes) by way of, for example, a laboratory analyzer (or other suitable analyzers) may be required if the comparator device is not labeled for use with the sample type of the higher-order materials. The "suitable calibrator" is any material with sufficiently well-characterized assigned target concentrations, e.g., secondary certified reference materials as defined by ISO 17511. Light grey boxes and dotted lines indicate parts of the traceability chain that are outside of the purview of bias estimation in a CGM performance study. Device types in the columns "measurement procedure/ device" are based on what is often used in CGM performance studies [12]. Note that the step between the comparator and the CGM systems is missing because CGM systems measure glucose concentrations in interstitial fluid, where comparator or reference measurements are currently not feasible. Similar traceability chains can be established for other sample origins.

Recommendations for the assessment and correction of bias

Bias towards the higher-order method's measurement results or assigned concentrations of the higher-order materials is used to verify the C-APS. Due to potential variation of bias over time, measurements with higher-order methods or materials should be performed sufficiently often, e.g., on every in-clinic day of a CGM performance study. They should cover the relevant glucose concentration range. Calibrator or reagent batch changes should be scheduled outside of the performance study's timeframe; otherwise, bias should be assessed using a sufficient amount of data for each combination of calibrator and reagent batch. The C-APS requirement for bias is defined as bias ≤2.1% [23]. Retrospective correction of comparator bias (sometimes also called "recalibration") should be used if the bias criterion is not met [20]. The IFCC WG-CGM recommends retrospective correction even if the bias criterion is met to further minimize differences between individual studies and between individual devices, if more than one comparator device is used in a study (e.g., in a multi-center study).

Assessment of measurement imprecision

Imprecision can, for example, be measured as the coefficient of variation (CV) of replicates from the same sample and pooled across all samples when there is no device drift, there are no changes in the device (like reagent or calibrator batch changes), and when the samples do not contain varying levels of interfering substances. In contrast to bias, it is not possible to minimize CV retrospectively. Imprecision of a potential comparator should, therefore, be assessed beforehand to minimize the risk of using an inadequate comparator. Imprecision of the comparator in a CGM performance study can be estimated from measuring replicates, e.g., duplicates, in each participant sample, and the average of these replicates can be used as comparator value. For a series of duplicate measurements, calculation of SD or CV is outlined using the "Dahlberg" formula by Kallner and Theodorsson [27]. The Dahlberg formula calculates the repeatability within series. Repeatability is different from the reproducibility between series, which is obtained by repeated measurements over many days of the same sample. Staying with the example of duplicate measurements, the effective imprecision of the comparator results when using averages of these duplicates can then be estimated from the standard error of the mean, i.e., CV is divided by the square root of the number of replicates. For duplicate measurements, the effective imprecision, expressed as effective CV, can be estimated as

$$CV_{eff} = \sqrt{\frac{2 \times \sum_{i=1}^{N}\left(\frac{x_{i,1}-x_{i,2}}{x_{i,1}+x_{i,2}}\right)^2}{2N}}$$

where $x_{i,1}$ and $x_{i,2}$ are the measurement results of the first and second replicate, respectively, and N is the total number of duplicates.

Imprecision can also be estimated from measurements with higher-order materials. If a split-sample comparison with a higher-order method is performed, multiple replicates are required.

Recommendations for the assessment of imprecision

The IFCC WG-CGM recommends that participant samples in the CGM performance study are measured in at least duplicates to allow estimation of imprecision as well as outlier detection [20,27]. The average of these (at least) duplicates should be used as comparator value to reduce the impact of imprecision. The C-APS requirement for imprecision is CV ≤2.4% [23]. Ideally, imprecision is characterized in advance to minimize the risk of inadequate imprecision in a CGM performance study.

Practical application of analytical performance specifications for comparators

Recently, a CGM performance study (German Clinical Trials Register, ID DRKS 00033697) was conducted. Here, both capillary and venous blood samples were collected for comparator measurements during frequent sampling periods, i.e., 7-hour periods where blood was sampled every 15±5 min. Cobas Integra 400 plus (INT; Roche Diagnostics GmbH, Mannheim, Germany) and, in total, three different YSI 2300 STAT Plus (YSI; YSI Incorporated, Yellow Springs, OH, USA) devices were used to obtain venous comparator data, and Contour Next (CNX; Ascensia Diabetes Care Holdings AG, Basel, Switzerland) was used to obtain capillary comparator data. Additional capillary measurements were performed twice per frequent sampling period with YSI and INT. Results from YSI devices were pooled since measurements from any given sample were only performed on one YSI device. Multiple YSI devices were used for feasibility reasons, as there were more venous samples on each in-clinic day than a single YSI device could have processed.

To assess and to allow for retrospective correction of bias of INT and YSI, 184 participant samples were sent in a split-sample approach for measurements with a spectrophotometric hexokinase-based method (Cobas pro [CBP]; Roche Diagnostics GmbH). The CBP was calibrated with the manufacturer's master calibrator, thus being of higher metrological order than INT and YSI, which were calibrated with marketed, i.e., end user, product calibrators. CBP's expanded measurement uncertainty (k=2) was reported as 3.5%, with negligible bias when compared regularly against IDMS. Bias of CNX was assessed based on the capillary measurements with INT and YSI and retrospectively corrected based on corrected INT and YSI results, respectively. There were slight differences in CNX measurement results depending on whether CNX was corrected retrospectively based on INT or YSI results. However, these were acceptably small and CNX's

bias was in both cases substantially lower after retrospective correction. If venous samples had not been collected in this study, then CBP could have been used as easily with capillary samples so that the traceability chain of Figure 1A would apply. Imprecision was estimated from the duplicate glucose measurements for all comparators [20,27]. Compliance with C-APS is shown in Table 2.

Table 2: Adherence with analytical performance specifications for comparators (C-APS).

Sample	Bias			Imprecision		C-APS met[2]	
Device	n	original	Corrected[1]	n	CV	original	final
Venous							
INT	184	+3.0%	+0.0%	2059	0.3%	No	Yes
YSI 1	64	-2.3%	+0.1%	706	0.2%	No	Yes
YSI 2	55	-1.5%	+0.0%	630	0.3%	Yes	Yes
YSI 3	65	-1.7%	+0.0%	724	0.6%	Yes	Yes
Capillary							
CNX (vs. INT)	142	+2.8%	+0.1%	2061	1.6%	No	Yes
CNX (vs. YSI)	142	+3.8%	+0.2%	2061	1.6%	No	Yes

As bias was corrected retrospectively, compliance with C-APS is given before and after retrospective correction.

CGM: continuous glucose monitoring system, FDA: Food and Drug Administration, BGMS: blood glucose monitoring system

CV: coefficient of variation, INT: Cobas Integra 400 plus, YSI: YSI 2300 STAT Plus, CNX: Contour Next

[1]For venous samples on INT and the three YSIs, bias was retrospectively corrected based on a split-sample approach with a higher-order method. For capillary CNX samples, bias correction was implemented using a split-sample approach using corrected capillary INT or YSI results (as indicated in the table).

[2]"Original" indicates application of C-APS to original bias and imprecision, whereas "final" indicates application of C-APS to retrospectively corrected bias and imprecision.

To illustrate the impact of bias on observed CGM performance, MARD results, 20%/20 agreement rates (ARs), i.e., percentages of CGM readings found within ±20% of paired comparator results ≥100 mg/dL and within ±20 mg/ dL of paired comparator results <100 mg/dL, and bias (mean relative difference) were calculated for one of the CGM systems investigated in the study mentioned above before and after recalibration (Table 3).

Table 3: Mean absolute relative difference (MARD), 20%/20 agreement rates (AR), i.e., percentages of CGM readings found within ±20% of paired comparator results ≥100 mg/dL and within ±20 mg/dL of paired comparator results <100 mg/dL, and bias (mean relative difference) for a current-generation CGM system.

Comparator	n	Before comparator bias correction			After comparator bias correction		
		MARD	20%/20 AR	Bias	MARD	20%/20 AR	Bias
INT (venous)	2001	9.5%	93.2%	+3.1%	10.5%	91.8%	+6.1%
YSI (venous)	2002	11.6%	88.9%	+8.4%	10.5%	91.2%	+6.3%
CNX (capillary)	2005	9.7%	92.4%	-1.1%	9.7% [a]	92.4% [a]	+1.6%[1]

AR: Agreement rate, MARD: mean absolute relative difference, INT: Cobas Integra 400 plus, YSI: YSI 2300 STAT Plus, CNX: Contour Next

[1]CNX results retrospectively corrected based on corrected INT results were used as comparator.

Before retrospective correction of comparator bias, there were substantial differences in the observed CGM performance depending on the venous comparator (INT vs. YSI). Correction of comparator bias eliminated the 2.1% MARD difference, reduced the approximately 4% difference in 20%/20 agreement rate, and reduced the difference in bias between the CGM system and the comparator from 5.3% to 0.2%. The differences in the original data could be viewed as relevant from both clinical and marketing perspectives despite being caused by the selection of the comparator and not being a property of the CGM system itself. When using the capillary CNX measurements as comparator, marked differences compared with venous results were found after retrospective correction of comparator bias, whose order of magnitude matches the capillary-to-venous differences in comparator glucose concentrations found in a recent study [13].

Summary and Outlook

In CGM performance studies, CGM readings are compared to values obtained with a comparator measurement procedure, whose description comprises comparator device, measurement method, and sample type. Different comparator methods have demonstrated biases of up to 8%, [18,19] different comparator devices of the same brand can systematically differ, on average, by more than 5%, [20,21] and concentrations in capillary and venous samples are physiologically different (often reported as 5 to 10% difference, sometimes more) [13].

The IFCC WG-CGM, therefore, recommends standardizing the comparator measurement procedure (Table 4). Previous recommendations were adopted [11]. Specifically, comparator measurements should be performed every 15 minutes for 6 to 8 hours during in-clinic sessions, and comparator data should exhibit certain characteristics, i.e., distribution of glucose concentrations and rates of change [11]. Furthermore, the use of capillary samples is recommended, mainly because CGM readings should align better with results from SMBG, which must be performed in some situations by people with diabetes:

1. some CGM systems require periodic manual calibration,
2. CGM systems typically require confirmation of very high or very low CGM readings,
3. SMBG is advised if symptoms do not match CGM readings, and
4. SMBG is the fallback option if CGM readings are not available, e.g., during warm-up or malfunction. For the comparator device and method, C-APS should be met; specifically, bias ≤2.1% and imprecision ≤2.4% [23] Bias between comparators can be reduced by retrospective correction of comparator values based on analysis with a method or materials of higher metrological order,[20] and correction is encouraged even if the bias criterion is met. Comparator samples should be measured in duplicate, and the averages should be used as the comparator result to improve imprecision of comparator results as well as to enable outlier detection.

A common reference and use of standardized procedures in CGM performance studies will minimize discrepancies and ensure alignment between readings across CGM systems.

Table 4: Recommendations of the IFCC WG-CGM regarding the comparator measurement procedure in CGM performance evaluations.

Category	Recommendation
Comparator sampling frequency	Every 15 minutes for 6 to 8 hours [11]
Comparator glucose concentration distribution	At least 7.5% of all values each covering the following situations: [11] a) BG < 70 mg/dL b) BG > 300 mg/dL, c) BG ≥ 70 mg/dL, RoC < -1 mg/dL/min, BG < 70 mg/dL within 30 min at current RoC, and d) BG ≤ 300 mg/dL, RoC > +1.5 mg/dL/min, BG > 250 mg/dL within 30 min at current RoC
Comparator sample type	Capillary samples
Comparator bias	Bias ≤2.1% across measuring range vs. method (or material) of higher metrological order,[23] retrospective correction encouraged even if criterion is met
Comparator imprecision	CV ≤2.4% across measuring range [23]
Comparator replicates	≥2, to assess and reduce imprecision [27] and to detect outliers in participant samples

BG: blood glucose concentration, CGM: continuous glucose monitoring, CV: coefficient of Variation, IFCC WG-CGM: Working Group on Continuous Glucose Monitoring of the International Federation of Clinical Chemistry and Laboratory Medicine, RoC: rate of change

Declaration of Conflicts of Interest
The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article. G.F. is general manager and medical director of the Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm; Germany (IfDT), which carries out clinical studies on its own initiative and on behalf of various companies. G.F./IfDT have received speakers' honoraria or consulting fees in the last 3 years Abbott, Berlin Chemie, Boydsense, Dexcom, Glucoset, i-SENS, Lilly, Menarini, Novo Nordisk, Perfood, Pharmasens, Roche, Sinocare, Terumo, Ypsomed.
S.P. and M.E. are employees of IfDT.
E.E.B. has no disclosures.
M.F. received lecture fees from Menarini.
R.H. – Independent medical & scientific consultant, former employee of Roche Diabetes Care
J.J. has been a lecturer/member of the scientific advisory boards at the following companies: Abbott, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Medtronic, Nordic InfuCare, NovoNordisk and Sanofi.
D.C.K. is a consultant for Afon, Better Therapeutics, Integrity, Lifecare, Nevro, Novo, Samsung, and Thirdwayv. K.M. has nothing to disclose.
J.H.N. has received research support from Abbott.
J.P. – Advisory panel for ROCHE Diabetes Care and Abbott. Speaker fees from Insulet and Dexcom.
E.S. is supported by grants from the US National Institutes of Health (NIH) and has received donated materials related to NIH-supported research from Abbott Diabetes Care, Roche Diagnostics, Siemens Diagnostics, Ortho Clinical Diagnostics, Abbott Diagnostics, Asahi Kasei Pharma Corp, GlycoMark Corp
N.K.T. is a consultant for Roche Diagnostics and Roche Molecular Systems, received honoraria from Nova Biomedical and Thermo Fisher, and Chair (2024-2026), Critical and Point-of-Care Testing (CPOCT) Division, Association for Diagnostics and Laboratory Medicine (ADLM).
L.W. has nothing to disclose.
R.J.S. is chair of the Clinical Chemistry Department of Isala which carries out clinical studies, e.g., with medical devices for diabetes therapy on its own initiative and on behalf of various companies. R.J.S. has received speaker's honoraria or consulting fees in the last three years from Roche and Menarini.

Abbreviations

AR	Agreement rate
BGMS	Blood glucose monitoring system
C-APS	Analytical performance specifications for comparators
CBP	Cobas pro
CGM	Continuous glucose monitoring
CLSI	Clinical and Laboratory Standards Institute
CNX	Contour Next
CV	Coefficient of variation
FDA	Food and Drug Administration
IDMS	Isotope dilution-mass spectrometry
IFCC WG-CGM	Working Group on Continuous Glucose Monitoring of the International Federation of Clinical Chemistry and Laboratory Medicine
INT	Cobas Integra 400 plus
MARD	Mean absolute relative difference
POC	Point-of-care
SD	Standard deviation
SMBG	Self-monitoring of blood glucose
YSI	YSI 300 STAT Plus

References
1. Teo E, Hassan N, Tam W, Koh S. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. Diabetologia. 2022;65(4):604-619. doi: 10.1007/s00125-021-05648-4.
2. Jancev M, Vissers TACM, Visseren FLJ, van Bon AC, Serné EH, DeVries JH, de Valk HW, van Sloten TT. Continuous glucose monitoring in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetologia. 2024;67(5):798-810. doi: 10.1007/s00125-024-06107-6.
3. Kumagai R, Muramatsu A, Fujii M, Katakura Y, Ito K, Fujie K, Nakata Y, Hashimoto K, Yagyu H. Comparison of glucose monitoring between Freestyle Libre Pro and iPro2 in patients with diabetes mellitus. Journal of diabetes investigation. 2019;10(3):851-856. doi: 10.1111/jdi.12970.
4. Michalak A, Pagacz K, Mlynarski W, Szadkowska A, Fendler W. Discrepancies between methods of continuous glucose

monitoring in key metrics of glucose control in children with type 1 diabetes. Pediatr Diabetes. 2019;20(5):604-612. doi: 10.1111/pedi.12854.

5. Freckmann G, Pleus S, Schauer S, Link M, Jendrike N, Waldenmaier D, Haug C, Stuhr A. Choice of Continuous Glucose Monitoring Systems May Affect Metrics: Clinically Relevant Differences in Times in Ranges. Exp Clin Endocrinol Diabetes. 2022;130(5):343-350. doi: 10.1055/a-1347-2550.

6. Pleus S, Kamecke U, Waldenmaier D, Link M, Zschornack E, Jendrike N, Haug C, Freckmann G. Time in Specific Glucose Ranges, Glucose Management Indicator, and Glycemic Variability: Impact of Continuous Glucose Monitoring (CGM) System Model and Sensor on CGM Metrics. J Diabetes Sci Technol. 2021;15(5):1104-1110. doi: 10.1177/1932296820931825.

7. Yeoh E, Png D, Khoo J, Chee YJ, Sharda P, Low S, Lim SC, Subramaniam T. A head-to-head comparison between Guardian Connect and FreeStyle Libre systems and an evaluation of user acceptability of sensors in patients with type 1 diabetes. Diabetes Metab Res Rev. 2022;38(7):e3560. doi: 10.1002/dmrr.3560.

8. Zhou Y, Mai X, Deng H, Yang D, Zheng M, Huang B, Xu L, Weng J, Xu W, Yan J. Discrepancies in glycemic metrics derived from different continuous glucose monitoring systems in adult patients with type 1 diabetes mellitus. J Diabetes. 2022;14(7):476-484. doi: 10.1111/1753-0407.13296.

9. Avari P, Tang W, Jugnee N, Hersi I, Al-Balah A, Tan T, Frankel AH, Oliver N, Reddy M. The Accuracy of Continuous Glucose Sensors in People with Diabetes Undergoing Hemodialysis (ALPHA Study). Diabetes Technol Ther. 2023;25(7):447-456. doi: 10.1089/dia.2023.0013.

10. Hanson K, Kipnes M, Tran H. Comparison of Point Accuracy Between Two Widely Used Continuous Glucose Monitoring Systems. J Diabetes Sci Technol. 2024;18(3):598-607. doi: 10.1177/19322968231225676.

11. Eichenlaub M, Pleus S, Rothenbühler M, Bailey TS, Bally L, Brazg R, Bruttomesso D, Diem P, Eriksson Boija E, Fokkert M, Haug C, Hinzmann R, Jendle J, Klonoff DC, Mader JK, Makris K, Moser O, Nichols JH, Nørgaard K, Pemberton J, Selvin E, Spanou L, Thomas A, Tran NK, Witthauer L, Slingerland RJ, Freckmann G. Comparator Data Characteristics and Testing Procedures for the Clinical Performance Evaluation of Continuous Glucose Monitoring Systems. Diabetes Technol Ther. 2024;26(4):263-275. doi: 10.1089/dia.2023.0465.

12. Freckmann G, Eichenlaub M, Waldenmaier D, Pleus S, Wehrstedt S, Haug C, Witthauer L, Jendle J, Hinzmann R, Thomas A, Eriksson Boija E, Makris K, Diem P, Tran N, Klonoff DC, Nichols JH, Slingerland RJ. Clinical Performance Evaluation of Continuous Glucose Monitoring Systems: A Scoping Review and Recommendations for Reporting. J Diabetes Sci Technol. 2023;17(6):1506-1526. doi: 10.1177/19322968231190941.

13. Pleus S, Link M, Hinzmann R, Öter S, Eichenlaub M, Baumstark A, Haug C, Freckmann G. Effect of Arterialization on Venous Blood Glucose Concentrations and Implications for Observed Continuous Glucose Monitoring Accuracy. Diabetes Technol Ther. 2024;26(4):238-245. doi: 10.1089/dia.2023.0489.

14. Clinical and Laboratory Standards Institute. POCT05 - Performance metrics for continuous interstitial glucose monitoring - Second edition. Wayne, PA: Clinical and Laboratory Standards Institute; 2020.

15. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, Kulpmann WR, Larsson L, Lewenstam A, Maas AH, Mager G, Naskalski JW, Okorodudu AO. Approved IFCC recommendation on reporting results for blood glucose: International Federation of Clinical Chemistry and Laboratory Medicine Scientific Division, Working Group on Selective Electrodes and Point-of-Care Testing (IFCC-SD-WG-SEPOCT). Clin Chem Lab Med. 2006;44(12):1486-1490. doi: 10.1515/CCLM.2006.275.

16. Brunkhorst FM, Wahl HG. Blood glucose measurements in the critically ill: more than just a blood draw. Crit Care. 2006;10(6):178. doi: 10.1186/cc5110.

17. Astles JR, Sedor FA, Toffaletti JG. Evaluation of the YSI 2300 glucose analyzer: algorithm-corrected results are accurate and specific. Clin Biochem. 1996;29(1):27-31. doi: 10.1016/0009-9120(95)02010-1.

18. Twomey PJ. Plasma glucose measurement with the Yellow Springs Glucose 2300 STAT and the Olympus AU640. Journal of clinical pathology. 2004;57(7):752-754. doi: 10.1136/jcp.2003.013417.

19. Freckmann G, Baumstark A, Jendrike N, Mende J, Schauer S, Link M, Pleus S, Haug C. Impact of Two Different Reference Measurement Procedures on Apparent System Accuracy of 18 CE-Marked Current-Generation Blood Glucose Monitoring Systems. J Diabetes Sci Technol. 2022;16(5):1076-1088. doi: 10.1177/1932296820948873.

20. Pleus S, Eichenlaub M, Gerber T, Eriksson Boija E, Makris K, Haug C, Freckmann G. Improving the Bias of Comparator Methods in Analytical Performance Assessments Through Recalibration. J Diabetes Sci Technol. 2024;18(3):686-694. doi: 10.1177/19322968221133107.

21. Bailey TS, Klaff LJ, Wallace JF, Greene C, Pardo S, Harrison B, Simmons DA. Fundamental Importance of Reference Glucose Analyzer Accuracy for Evaluating the Performance of Blood Glucose Monitoring Systems (BGMSs). J Diabetes Sci Technol. 2016;10(4):872-875. doi: 10.1177/1932296816634356.

22. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, Jonker N, Simon M, Braga F, Perich C, Boned B, Marques-Garcia F, Carobene A, Aslan B, Sezer E, Bartlett WA, Sandberg S. The EFLM Biological Variation Database [Accessed 28/01/2025]. Available from: https://biologicalvariation.eu/.

23. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Lernmark Å, Metzger BE, Nathan DM, Kirkman MS. Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. Diabetes Care. 2023;46(10):e151-e99. doi: 10.2337/dci23-0036.

24. Kölle J, Eichenlaub M, Mende J, Link M, Vetter B, Safary E, Pleus S, Haug C, Freckmann G. Performance Assessment of Three Continuous Glucose Monitoring Systems in Adults With Type 1 Diabetes. J Diabetes Sci Technol. 2024;18(6):1424-1432. doi: 10.1177/1932296823159657.

25. Han J, Heinemann L, Ginsberg BH, Alva S, Appel M, Bess S, Chen KY, Freckmann G, Harris DR, Hartwig M, Hinzmann R, Kerr D, Krouwer J, Morrow L, Nichols J, Pfutzner A, Pleus S, Rice M, Sacks DB, Schlueter K, Vesper HW, Klonoff DC. The YSI 2300 Analyzer Replacement Meeting Report. J Diabetes Sci Technol. 2020;14(3):679-686. doi: 10.1177/1932296820911471.

26. Mader JK, Waldenmaier D, Mueller-Hoffmann W, Mueller K, Angstmann M, Vogt G, Rieger CC, Eichenlaub M, Forst T, Freckmann G. Performance of a Novel Continuous Glucose Monitoring Device in People With Diabetes. J Diabetes Sci Technol. 2024;18(5):1044-1051. doi: 10.1177/19322968241267774.

27. Kallner A, Theodorsson E. Repeatability imprecision from analysis of duplicates of patient samples and control materials. Scandinavian journal of clinical and laboratory investigation. 2020;80(3):210-214. doi: 10.1080/00365513.2019.1710243.