Title

Clinical Performance Evaluation of Continuous Glucose Monitoring Systems: A Scoping Review and Recommendations for Reporting.

Journal

DST, 17(6)

Authors

Freckmann, Guido

Eichenlaub, Manuel

Waldenmaier, Delia

et al.

Publication Date

2023-11-01

DOI

10.1177/19322968231190941

Peer reviewed

Journal of Diabetes Science and Technology
2023, Vol. 17(6) 1506–1526

DOI: 10.1177/19322968231190941
journals.sagepub.com/home/dst
Sage

Clinical Performance Evaluation of Continuous Glucose Monitoring Systems: A Scoping Review and Recommendations for Reporting

Guido Freckmann, MD[1,2], Manuel Eichenlaub, PhD[2], Delia Waldenmaier, PhD[2], Stefan Pleus, PhD[2], Stephanie Wehrstedt, PhD[2], Cornelia Haug, MD[2], Lilian Witthauer, PhD[3,4], Johan Jendle, MD, PhD[1,5], Rolf Hinzmann, MD, PhD[1,6], Andreas Thomas, PhD[1,7], Elisabet Eriksson Boija, PhD[1,8], Konstantinos Makris, PhD[1,9], Peter Diem, MD[1,10], Nam Tran, PhD[1,11], David C. Klonoff, MD, FACP, FRCP (Edin), Fellow AIMBE[1,12], James H. Nichols, PhD[1,13], and Robbert J. Slingerland, PhD[1,14]; on behalf of the Working Group on Continuous Glucose Monitoring of the IFCC Scientific Division

Abstract

The use of different approaches for design and results presentation of studies for the clinical performance evaluation of continuous glucose monitoring (CGM) systems has long been recognized as a major challenge in comparing their results. However, a comprehensive characterization of the variability in study designs is currently unavailable. This article presents a scoping review of clinical CGM performance evaluations published between 2002 and 2022. Specifically, this review quantifies the prevalence of numerous options associated with various aspects of study design, including subject population, comparator (reference) method selection, testing procedures, and statistical accuracy evaluation. We found that there is a large variability in nearly all of those aspects and, in particular, in the characteristics of the comparator measurements. Furthermore, these characteristics as well as other crucial aspects of study design are often not reported in sufficient detail to allow an informed interpretation of study results. We therefore provide recommendations for reporting the general study design, CGM system use, comparator measurement approach, testing procedures, and data analysis/statistical performance evaluation. Additionally, this review aims to serve as a foundation for the development of a standardized CGM performance evaluation procedure, thereby supporting the goals and objectives of the Working Group on CGM established by the Scientific Division of the International Federation of Clinical Chemistry and Laboratory Medicine.

Keywords

continuous glucose monitoring, clinical performance evaluation, study design, accuracy

Introduction

Systems for continuous glucose monitoring (CGM) have been available for more than 20 years, but significant advances have been made in the recent past and CGM performance has reached a level that allows non-adjunctive use for clinical decision-making.[1] There are now several CGM systems available for routine clinical use, and CGM-derived therapy parameters are part of national and international guidelines for the therapy of diabetes mellitus.[2,3] Furthermore, CGM data are routinely used to assess outcomes in clinical trials of diabetes treatments. As a result, the performance of CGM systems, in particular accuracy, has gained renewed importance, resulting in a surge of published studies examining CGM performance in the recent past.

The performance of a CGM system can include a wide range of aspects that characterize the properties of the CGM system. In this article, we will focus on the process of a clinical performance evaluation of CGM systems and its endpoints,

Table 1. Description of the Endpoints of a Clinical Performance Evaluation of CGM Systems Discussed in the Context of This Article, Based on the POCT05 Guideline.[4]

Endpoints of a clinical CGM performance evaluation
Accuracy
• Analytical point accuracy, ie, the characterization of deviations between CGM values and blood glucose concentrations measured with a suitable comparator at a single point in time, including bias, precision, and sensor-to-sensor variability
• Clinical point accuracy, ie, the clinical interpretation of the impact of the analytical point accuracy
• Trend accuracy, ie, the characterization of deviations between the glucose concentration rate of change (RoC) indicated by the CGM system and blood glucose concentrations measured with a suitable comparator, including a clinical interpretation of those deviations
Stability
• Sensor stability, ie, the accuracy with respect to the sensor lifetime
• Calibration stability, ie, the accuracy with respect to the time after calibration
Alert reliability
• Threshold alert reliability, ie, the ability to correctly alert the user when glucose concentrations cross predefined hypo- and hyperglycemic thresholds, including the measurement range limits of the CGM system
• Predictive alert reliability, ie, the ability to correctly predict the crossing of predefined hypo- and hyperglycemic thresholds
Technical reliability
• Sensor survival, ie, the ability of a CGM sensor to correctly function until the end of its specified use lifetime
• Data availability, ie, the ability of a CGM system to provide the expected number of glucose measurements without interruptions
• Description of device deficiencies, ie, malfunctions and use errors and their potential to cause an adverse event
• Lag time between CGM and comparator data
Adverse device effects
• Occurrence of adverse events related to or caused by the examined CGM systems
User satisfaction
• Experience of the users with the CGM system, eg, in terms of ease-of-use, functionality, reliability, and wear comfort

Abbreviation: CGM, continuous glucose monitoring.

described in Table 1. Note that these endpoints go beyond the investigation of accuracy and encompass other aspects such as alarm and technical reliability; however, a detailed description of how to quantify these different endpoints lies beyond the scope of this article. Aspects determined by pre-clinical testing are also not covered in this article. These include biocompatibility, electrical safety, and cyber security as well as the effect of interfering substances (both pre-clinical and clinical). Additionally, the evaluation of the clinical benefit and usability is not covered.

When designing a study for clinical CGM performance evaluation, there exist various elements, such as the subject population, the testing protocol, the selection of comparator samples and methods, and the definition of statistical analysis, that all provide several options that must be carefully considered. Here, it has been established that the choice of study design and evaluation procedures affect the observed accuracy leading to a range of varying accuracy results reported for each CGM system.[5] This indicates that there is no single level of performance for any given CGM system, but potentially as

[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]Diabetes Center Berne, Bern, Switzerland
[4]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital Bern, Bern University Hospital and University of Bern, Bern, Switzerland
[5]School of Medical Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden
[6]Roche Diabetes Care GmbH, Mannheim, Germany
[7]Pirna, Germany
[8]Equalis AB, Uppsala, Sweden
[9]Clinical Biochemistry Department, KAT General Hospital, Athens, Greece
[10]Endokrinologie Diabetologie Bern, Bern, Switzerland
[11]Department of Pathology and Laboratory Medicine, University of California Davis Health, Sacramento, CA, USA
[12]Diabetes Research Institute, Mills-Peninsula Medical Center, San Mateo, CA, USA
[13]Vanderbilt University Medical Center, Nashville, TN, USA
[14]Department of Clinical Chemistry, Isala Clinics, Zwolle, the Netherlands

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.
Email: manuel.eichenlaub@idt-ulm.de

many levels as there are performance studies with any specific CGM system. Consequently, it is only possible to compare different systems when they are tested in the same study or if they are tested according to a standardized protocol.[6,7]

This issue of limited comparability persists until today, despite early calls for standardized clinical performance evaluations in 2007,[8] which resulted in the Clinical and Laboratory Standards Institute guideline POCT05, providing specifications for study and evaluation procedures. This guideline was introduced in 2008[9] and revised in 2020.[4] While the guidance document provides a comprehensive overview of the statistical evaluation of the data, it does not provide specific guidance regarding some important details of study design, in particular the testing procedures. Therefore, it is rarely used as a complete evaluation protocol, but rather individual elements of the guideline can be found in different study protocols. In this review, we found that only 11 studies published since 2010 cite the POCT05 guideline (first or second edition), suggesting a limited adoption.

A more concise step toward standardization was made by the U.S. Food and Drug Administration (FDA) in 2018 with the definition of acceptance criteria for "integrated" CGM (iCGM) systems.[10,11] However, the FDA requirements are also vague regarding the design of the clinical studies and only state that "clinical data must be obtained from a clinical study designed to fully represent the performance of the device throughout the intended use population and throughout the measuring range of the device" (special control 1 [iii]).[10] To obtain a Conformité Européenne (CE) mark, which is necessary for product launch in most European countries, no dedicated requirements for the clinical CGM performance evaluations and its outcomes have been published. An excellent and detailed review on the regulatory approval of CGM systems in Europe, the United States, and Australia has recently been published by Pemberton et al.[12]

Existing review articles on CGM accuracy found that the comparability of results is limited because of different procedures. However, these reviews have not characterized the different aspects of study design and performance presentation in detail.[5,12-15] This scoping literature review therefore offers a comprehensive overview of the relevant elements of clinical CGM performance evaluations published between 2002 and 2022. In addition, we provide recommendations on how clinical CGM performance studies should be presented in original articles, beyond general reporting recommendations such as CONSORT or STROBE.[16] This article is intended to support the efforts of the Working Group on CGM (WG-CGM) of the Scientific Division of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) in their aim to define a standardized procedure for the clinical performance evaluation of CGM systems.[7,17] A further goal of this review is to support journal editors, reviewers, and regulators who are judging the quality of reports of clinical CGM performance evaluations.

The article is structured as follows. First, the methodology of the literature search and selection process are presented. Next, the review examines several aspects of clinical CGM performance studies, including general study design, use of CGM systems, comparator measurements, testing protocol, and data analysis and statistical performance evaluation. Each aspect is thoroughly reviewed and discussed, and recommendations for reporting are provided. A table with all studies included in this review together with their most relevant features and results is provided in the Supplemental Material.

Methodology of Literature Search

This scoping review is based on a systematic literature search conducted on the MEDLINE database. Therefore, this review only includes study reports published in scientific journals and excludes reports published outside of these sources, such as those from regulatory agencies or manufacturer white papers. Additionally, we should explicitly state that, because of the abundance of studies, the authors of the study were not contacted for additional data. Therefore, only information published in the main text or supplemental material of the peer-reviewed manuscripts was considered.

The search term ("CGM" OR "continuous glucose sensor" OR "continuous glucose monitoring" OR "continuous subcutaneous glucose monitoring" OR ["flash" AND "glucose"]) AND ("accuracy"[Title] OR "performance"[Title] OR "evaluation"[Title] OR "comparison"[Title]) NOT "review"[Title] NOT review[Filter] NOT (Animal[Filter]) was defined, using a period of search from 2002 to 2022. Titles and abstracts of all retrieved articles were screened. From the first stage of the selection process, we excluded manuscripts not available in the English language or if they did not contain a CGM performance evaluation. In the second selection stage, the remaining studies were inspected in more detail and excluded if one or more of the following criteria were met:

1. The manuscript was in a short format, such as a brief report or letter to the editor.
2. The manuscript described a study investigating CGM accuracy in a specific subject population aside from people with diabetes under routine care. We thus excluded studies conducted exclusively on infants, subjects with comorbidities, pregnant women, and subjects without diabetes. Furthermore, we excluded studies with patients in intensive care and patients undergoing surgery.
3. The manuscript described a study evaluating a CGM system not designed to measure glucose in the interstitial fluid (ISF).
4. The manuscript described a feasibility study intended for proof-of-concept testing including studies conducted for calibration algorithm development.

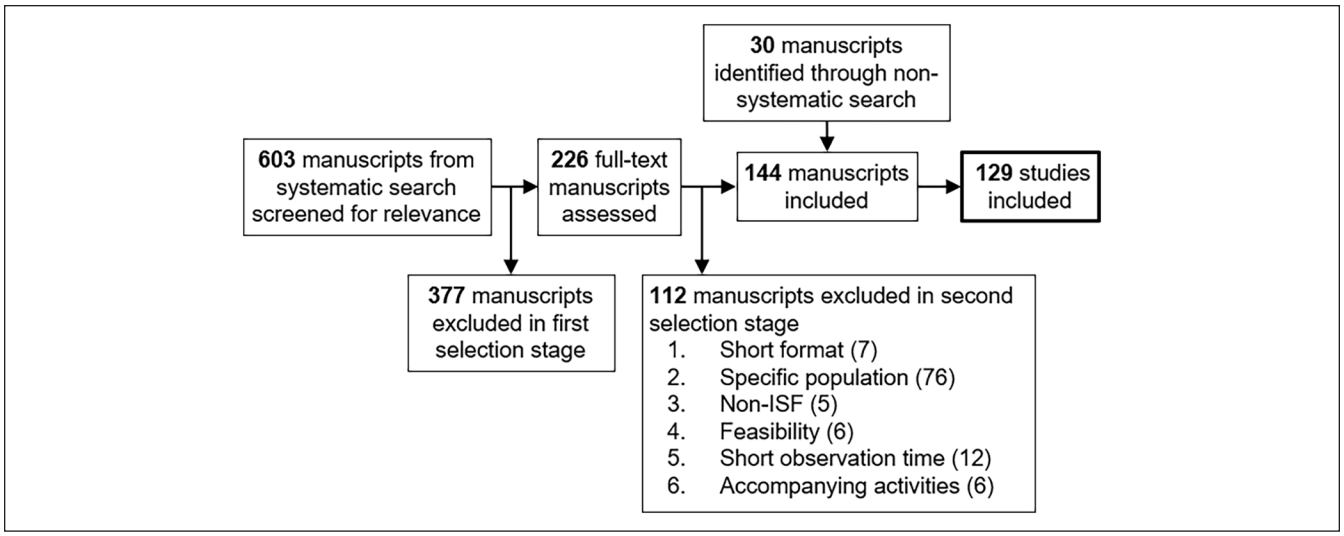

Figure 1. Schematic depiction of the study selection process.

5. The manuscript described a study with observation periods for comparator measurement shorter than two hours.
6. The manuscript described a study investigating CGM accuracy under specific accompanying activities, eg, during ethanol consumption or glucagon injection.

Suitable manuscripts not found in the systematic search were also identified using the "similar articles" function and screening the references of included manuscripts (non-systematic search).

Studies included in the final review were systematically summarized by at least two researchers. If a single study was presented in multiple manuscripts, then these manuscripts were grouped for evaluation.

The process of article selection is depicted in Figure 1. The initial search yielded 603 results, of which 377 manuscripts were excluded in the first selection stage. The remaining 226 manuscripts were assessed and 112 additional manuscripts were excluded in the second selection stage, yielding 114 manuscripts. Another 30 manuscripts identified from the non-systematic search were included, yielding a total of 144 manuscripts.[18-161] This high number of manuscripts identified in the non-systematic search is the result of the inconsistent terminology for "continuous glucose monitoring system" in the titles, where various slightly varying synonyms, not covered by the systematic search term, were used. Additionally, the titles of several manuscripts only included the brand name of the tested devices. After summarizing manuscripts using data from the same study, a total of 129 studies were included in this scoping review.

The selected articles showed a trend toward an increasing number of publications over the years with a peak between 2017 and 2019 (Figure 2).

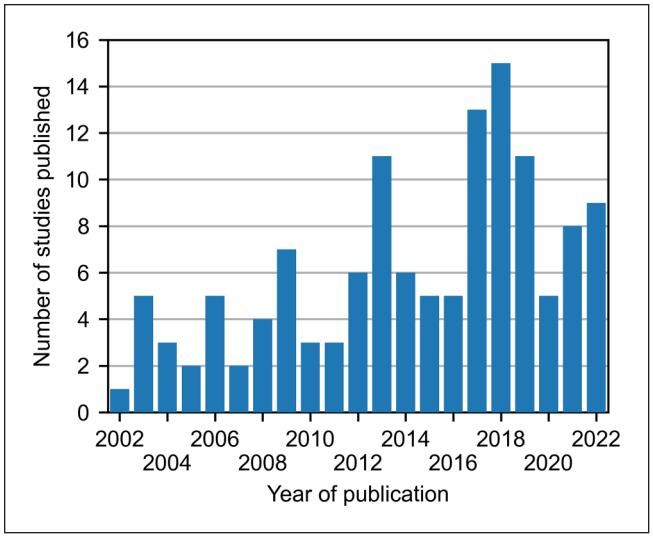

Figure 2. Distribution of the years of publication of the 129 performance studies included in this review.

General Study Design

Subject Population

The number of subjects included in a clinical CGM performance study is one of the most crucial aspects of subject population selection. Unlike for blood glucose monitoring systems (BGMS),[162,163] no requirements on sample size for market approval of CGM systems have been put forward by regulatory agencies. In this context, the review by Pemberton et al[12] found that the number of adult subjects enrolled in studies for CE marking of CGM systems prescribed in UK primary care ranged from 57 to 316.

The appropriate sample size depends on the specific purpose of the study, the expected accuracy of the system, and

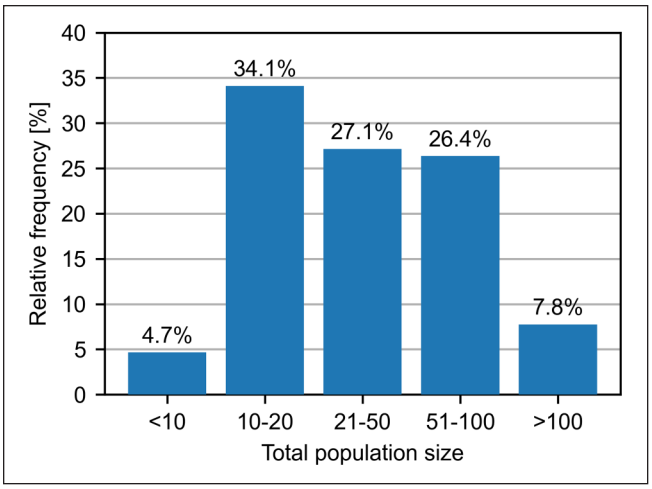

Figure 3. Distribution of the total number of subjects in the 129 studies included in this review.

potential minimum accuracy requirements. The studies included in this review had population sizes ranging from 6 to 318 subjects with a median of 30 and an interquartile interval between 16 and 63 subjects. A more detailed distribution of population sizes of the studies included in this review is shown in Figure 3, which demonstrates a large variability in sample size although feasibility studies, which typically have comparatively small sample sizes, were excluded. Despite being a crucial aspect of study design that should be reported, a sample size calculation or at least a justification of the chosen sample size was only provided in 14.7% of the articles.

Another important aspect of the selected population is the diabetes type of the subjects which depends on the intended use of the examined CGM system and affects the characteristics of the comparator data. Most of the studies (65.9%) only included subjects with type 1 diabetes, which can be explained by the fact that this facilitates the collection of comparator data containing a wide range of blood glucose concentrations. A total of 30.2% of studies included people with both type 1 and type 2 diabetes, where the percentage of people with type 1 diabetes from the entire population had a large variability ranging from 8% to 99%, with a median of 73.5%. The populations of the remaining 3.9% of studies consisted of only people with type 2 diabetes or a mixture of people with diabetes (type 1 and type 2) and without diabetes.

The next aspect to be reviewed is the subjects' age. Here, most studies (76.0%) included adults only, and the remaining studies included both adults and minors or minors only. This can be explained by the increased complexity and ethical challenges when performing clinical studies with minors. The influence of subject age on the testing protocol will be discussed in a later section. In total, 23.3%, 17.8%, and 7.8% of the reviewed studies included adolescents (age 12-17 years), children (age 6-11 years), and/or young children (age <6 years), respectively.

Finally, the median percentage of male subjects out of the total population was 54.5% with values ranging from 20.0% to 91.7%. However, 10.9% of studies did not report the subject sex, at all. In terms of ethnicity, only 24.8% reported the ethnicity of the subjects.

Number of Study Centers

Less than half (41.1%) of the studies included in this review had a multicenter design. As these studies allowed the recruitment and data collection at multiple study centers under the same protocol, the median population size of multicenter studies was 71 (interquartile interval 38-90), compared with 20 (interquartile interval 14-30) in monocenter studies. Because of potential systematic differences between glucose analyzers for comparator measurement, recently discussed by Pleus et al,[164] an important aspect to consider in multicenter studies is the harmonization of devices for comparator measurement. However, only 9.4% of multicenter studies using laboratory analyzers described any measures or procedures to ensure harmonization.[19-22,29,33,38,42,43] These studies were almost conducted by the Diabetes Research in Children Network, where blood samples from all study centers were transferred and analyzed in a central laboratory. In multicenter studies, we recommend reporting the approach for harmonization of comparator data as unharmonized comparator data could affect the observed performance.

Funding Source

A review from 2017 on the association between industry sponsorship and research outcome concluded that "sponsorship of drug and device studies by the manufacturing company leads to more favorable efficacy results and conclusions than sponsorship by other sources."[165] The role of sponsoring by CGM system manufacturers was therefore examined in this review, where it was found that 58.1% of studies were at least partially sponsored by a CGM system manufacturer. A study was considered manufacturer-sponsored if it was explicitly declared as fully or partially funded by a manufacturer of the tested CGM systems, or if the study devices of at least one CGM system were provided by the manufacturer free of charge. Furthermore, if studies were fully or partially funded by a manufacturer, it was often not clear to what extent study design, procedures, and data evaluation were influenced by the manufacturer. This emphasizes the need for detailed and transparent reporting of all study aspects so that potential biases in the study design can be identified. Furthermore, we recommend describing the influence of the funding source on the study planning and implementation process, as well as on the evaluation and interpretation of data. Examples of such a description were found in the articles by Kropff et al,[92] Ólafsdóttir et al,[111] and Piona et al.[122]

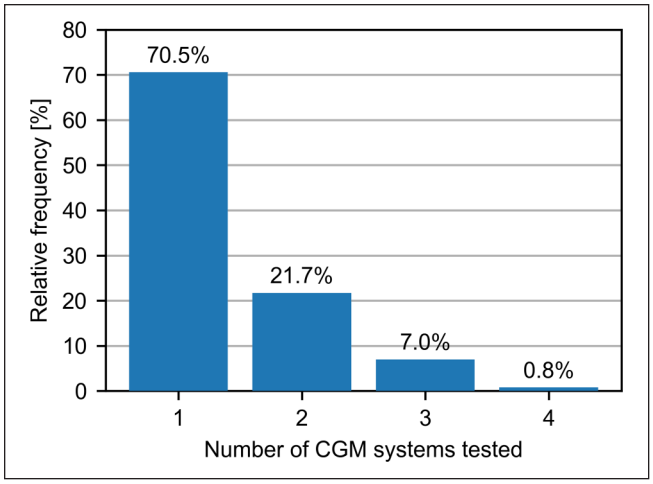

Figure 4. Distribution of the number of CGM systems that were tested simultaneously in the 129 studies included in this review.

Use of CGM Systems

The total number of tested CGM systems, calculated as the sum of the number of different systems tested in each study, was 178. However, of all reviewed studies, only 29.5% tested more than one, and up to four CGM systems in parallel (see Figure 4). These studies, often referred to as head-to-head studies, are of particularly high value because they are currently the only approach to genuinely evaluate and compare the performance of two or more different CGM systems, given that all factors influencing their performance are equal. Such studies have been performed with earlier-generation CGM systems, eg, Kovatchev et al,[44] Damiano et al,[69,81] and Freckmann et al.[70,117,132] However, to date, no head-to-head studies comparing the performance of most current CGM systems, including Dexcom G7, FreeStyle Libre 2 and 3, and Medtronic Guardian Sensor 4, have been published. Dexcom G6 was only examined in one study with other CGM systems (FreeStyle Libre and Medtronic Enlite 2 in combination with MiniMed 640G, systems not worn simultaneously) during a summer camp for children and adolescents with type 1 diabetes.[150]

In comparison with head-to-head designs, inserting multiple sensors of the same CGM system to the same subject is more common, because it is one of the few approaches to examine sensor precision. Furthermore, this approach can increase the number of data points for analysis without requiring more subjects. Here, 39.5% of studies used more than one sensor of at least one of the tested systems simultaneously. Depending on the study duration, multiple sensors of the same CGM system can also be worn consecutively, which is often the case in studies mainly conducted in a free-living setting.

In this context, the use of sensors from different manufacturing lots should be addressed. Here, only six studies[71,91,103,115,130,132] provided information on the number of sensor lots used, although the POCT05 guideline recommends testing on at least three lots.[4] This is important because it might be possible that sensor performance might change over time due to varying manufacturing conditions. We therefore also recommend that investigators provide information on the used manufacturing lots.

Out of the 178 different CGM systems tested, the principal manufacturers Abbott Diabetes Care, Dexcom, and Medtronic had similar percentages of 26.4%, 25.8%, and 29.8%, respectively, with the remaining 18.0% of tested CGM systems manufactured by other companies, eg, Senseonics or A. Menarini Diagnostics. In this context, we recommend reporting the source of the examined devices and sensors. In particular, in manufacturer-independent studies, it should be stated where the sensors and devices were purchased from (ie, directly from the manufacturer or through regular channels from the market).

The sensors of the studies included in this review were most commonly inserted on the abdomen, followed by the upper arm, and we recommend that studies clearly state whether the sensors were inserted by the subjects themselves, under the supervision of study personnel, or directly by the study personnel.

Calibration is a major topic concerning CGM accuracy even though factory-calibrated systems are becoming more prevalent. In most studies, the calibration was performed according to the instructions of the manufacturers, although in some cases this was not clearly reported. A few studies, eg, by Zueger et al[64] and Bailey et al,[80] examined the effect of additional and/or timing of calibrations on CGM accuracy.

Comparator Measurement Approach

The topic of comparator method selection and establishment of traceability in clinical CGM performance evaluations has been discussed in a previous article by the IFCC Working Group on CGM.[7]

CGM sensors are typically placed in the ISF and use algorithms to calculate blood glucose concentrations; however, the compartment (typically capillary or venous blood) for which CGM systems display glucose concentrations is often not specified by the manufacturer and not standardized. Because of this and the impossibility to perform comparator measurements in the ISF, three possible compartments for comparator glucose measurements were used in the reviewed studies: capillary, venous, and arterialized-venous, where venous blood is sampled from a specifically heated arm or hand. Here it should be emphasized that there are physiological differences between these compartments that can influence glucose concentrations and thus the observed performance of the CGM systems. The total number of used comparators, calculated as the sum of the number of different comparators used in each study, was 167. The percentages of comparators using capillary, venous, and arterialized-venous blood are shown in Figure 5.

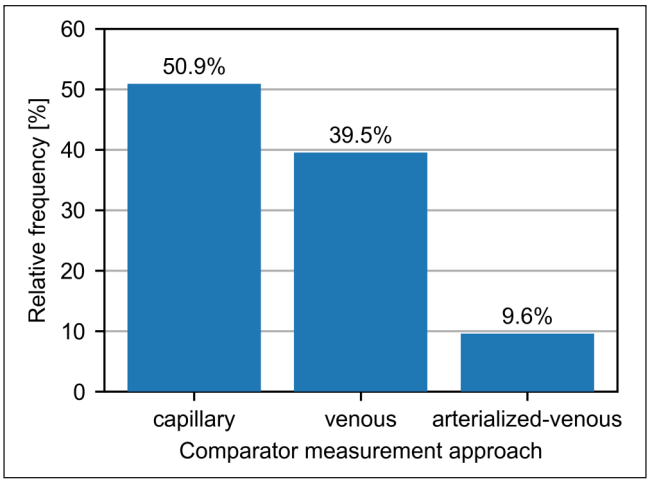

Figure 5. Relative frequencies of capillary, venous, and arterialized-venous sampling of the 167 comparators found in this review.

Capillary blood is typically sampled via finger-prick, which requires no preprocessing when glucose concentrations are measured with a BGMS and the measurements can be carried out in a free-living setting by the subjects themselves (the study setting is discussed in the next section). In this setting, capillary blood glucose concentrations have been used for therapy decisions before the advent of CGM and are also used for CGM system calibration. Even in non-adjunctive CGM systems, capillary measurements are still recommended when CGM values are implausible or temporarily unavailable. A disadvantage of using capillary comparator measurements is the repeated finger-pricks during phases of frequent sampling. Furthermore, a BGMS usually has a reduced measurement accuracy in comparison to laboratory-grade glucose analyzers, partially caused by pre-analytic errors, keeping in mind that any inaccuracy in the comparator measurements is fully attributed to the CGM system. In most studies where capillary comparator measurements were obtained, a BGMS was used (85.9%). Only few studies (14.1%) used laboratory-grade glucose analyzers with capillary blood, which can be more accurate than BGMS but also require a larger sample volume, thus impeding frequent sampling.

In almost all reviewed studies where venous and/or arterialized-venous blood was collected, sampling was limited to in-clinic visits, because obtaining these specimens requires venous access as well as pre-analytical processing, such as centrifugation to separate plasma (except the study by Francescato et al,[59] in which venous samples were collected at the subjects' homes). One advantage of venous sampling is that samples can be collected frequently without repeated skin penetration. Glucose concentrations are then measured with a laboratory analyzer, of which a Yellow Springs Instruments (YSI) analyzer based on the glucose-oxidase technology was used most often (65.9%). Laboratory analyzers from other manufacturers were, for example, based on the hexokinase enzyme. Here it should be mentioned that it has been shown that there are systematic differences between laboratory analyzers, which can affect the observed CGM accuracy.[164,166] We therefore recommend examining and reporting the compliance of all comparators with the recently proposed analytical performance specifications of comparators used in clinical CGM performance evaluations.[164] An example of such an examination can be found in a recent study by Kölle et al,[167] which was excluded from this review because it was published in 2023. We further recommend describing the sample compartment, sample processing, and comparator glucose analyzers in sufficient detail (eg, Guerci et al,[26] Francescato et al,[59] and Yan et al[144]).

Testing Procedures

In principle, comparator data can be collected either in an in-clinic or free-living setting, keeping in mind that study endpoints such as accuracy and alert reliability can only be evaluated during times of comparator data collection. The choice of setting affects the method for comparator measurement, discussed in the previous section, as well as the amount, frequency, and characteristics of comparator data. The free-living setting represents real-life use of the CGM systems and can provide comparator data (using a BGMS) on every day of the sensor lifetime because subjects are typically instructed to measure multiple times per day. In contrast, in-clinic visits are typically sporadic and allow testing under controlled conditions with frequent collection of comparator samples, use of laboratory analyzers for comparator measurement, and deliberate manipulation of glucose concentration dynamics under the supervision of a study physician. Considering only the setting for which data were analyzed and presented, this review found that most studies (55.0%) only used data collected in an in-clinic setting, whereas 21.7% of studies only evaluated data collected in a free-living setting. The remaining 23.3% of studies used data collected in both settings and the data were either pooled or analyzed separately.

In-Clinic Setting Protocol

The in-clinic setting protocol for the collection of comparator measurements can involve multiple separate visits of varying durations. Of the 101 studies that included in-clinic data collection, 46.5% only involved a single visit. Two, three, and four or more visits occurred in 23.8%, 15.8%, and 13.9% of studies, respectively. Naturally, the number of visits depends on the sensor lifetime. While early-generation sensors had shorter maximum lifetimes of two to three days that necessitated only one or two visits, lifetimes of more than a week found in newer-generation sensors require several visits or a stationary admission of subjects to examine accuracy over the complete sensor lifetime. In this context, it should be mentioned that CGM performance can be different

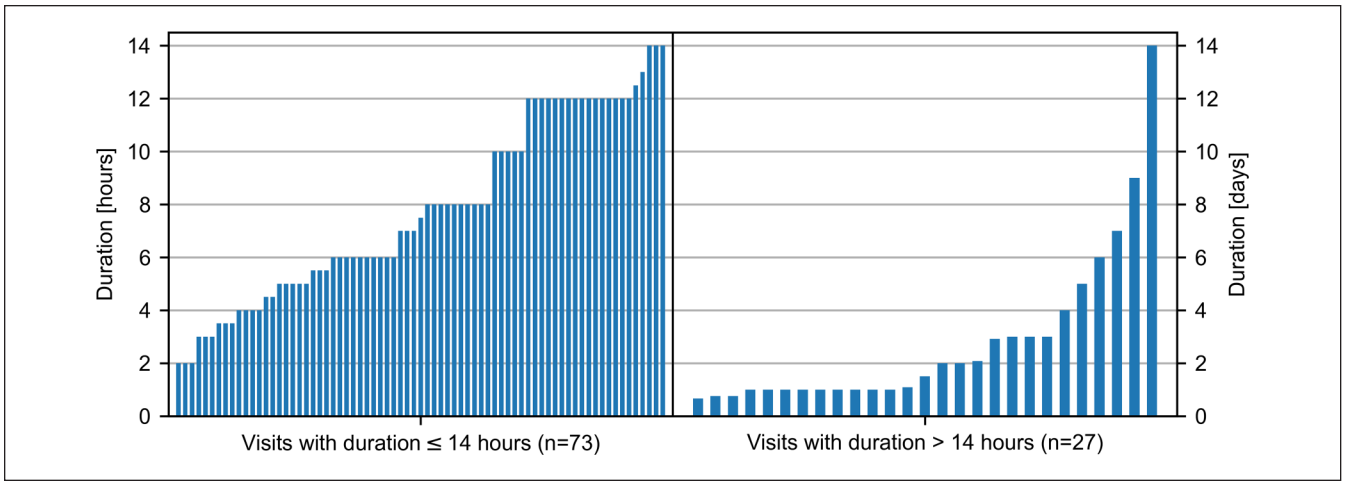

Figure 6. Duration of in-clinic visits of individual studies separated by visits shorter or equal to 14 hours in hours (left) or longer than 14 hours in days (right).

depending on the sensor wear time, with performance typically being reduced at the beginning and end of the sensor lifetime. The chosen distribution of in-clinic visits with respect to the day of sensor wear can thus have an impact on the observed performance.

Examining the duration of visits, it was found that 26.7% of studies with in-clinic settings included visits lasting for more than 14 hours, which typically requires stationary admission of the subjects and allows data collection during the night. Considering studies with in-clinic visits of 14 hours or less, the median duration was 7.5 hours with an interquartile interval between 5 and 12 hours. A detailed overview on the durations of in-clinic visits is provided in Figure 6.

Another important aspect of the in-clinic protocol is the schedule of comparator sample collection. Here it was found that 82.2% of studies with in-clinic visits had at least one period where comparator measurements were carried out every 20 minutes or more frequently. This threshold was chosen as it allows the estimation of blood glucose rate of change (RoC) and allowed the inclusion of the numerous studies that reported the sampling interval as 15±5 minutes. Furthermore, it was common to have a varying sampling schedule or to adapt the sampling interval depending on the glucose concentration. In particular, some of the sampling intervals were reported to be as short as five minutes. Here, some authors voiced their concern about the possible statistical interdependency of samples collected more frequently than every 15 minutes.[73,92,112] However, to the best of the authors' knowledge, a more thorough statistical examination of those concerns is currently missing. Another concern is the total blood volume extracted from the subjects, for which the POCT05 guideline provides recommendations.[4]

The last aspect to consider in this section is the deliberate manipulation of glucose concentrations. This was the case in 73.3% of studies that included in-clinic sessions, whereas

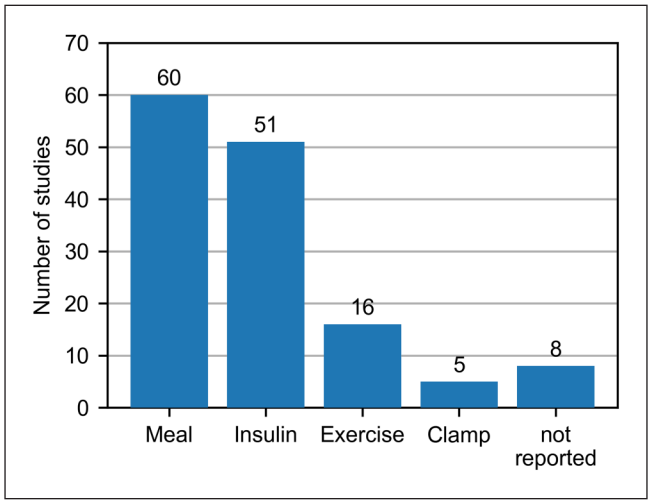

Figure 7. Number of studies employing various approaches for deliberate glucose manipulations.

18.8% explicitly stated that no deliberate manipulation took place and 7.9% provided no information. The predominant methods to manipulate glucose concentrations were the adaptation of insulin dosing and/or timing as well as providing specific meals, at least one of which occurred in over 90% of studies with in-clinic sessions and glucose manipulations. Less common was the use of exercise, possibly in combination with meal and insulin manipulation, occurring in 21.6% of those studies. The least common was the use of glucose clamping techniques, which was reported in only five studies. Here glucose concentrations were precisely regulated by external, intravenous infusion of glucose and insulin.[31,44,51,57,110] More details are provided in Figure 7. When CGM systems are tested with minors, it is common to reduce the number of visits and/or visit duration and to refrain from deliberate glucose manipulations.

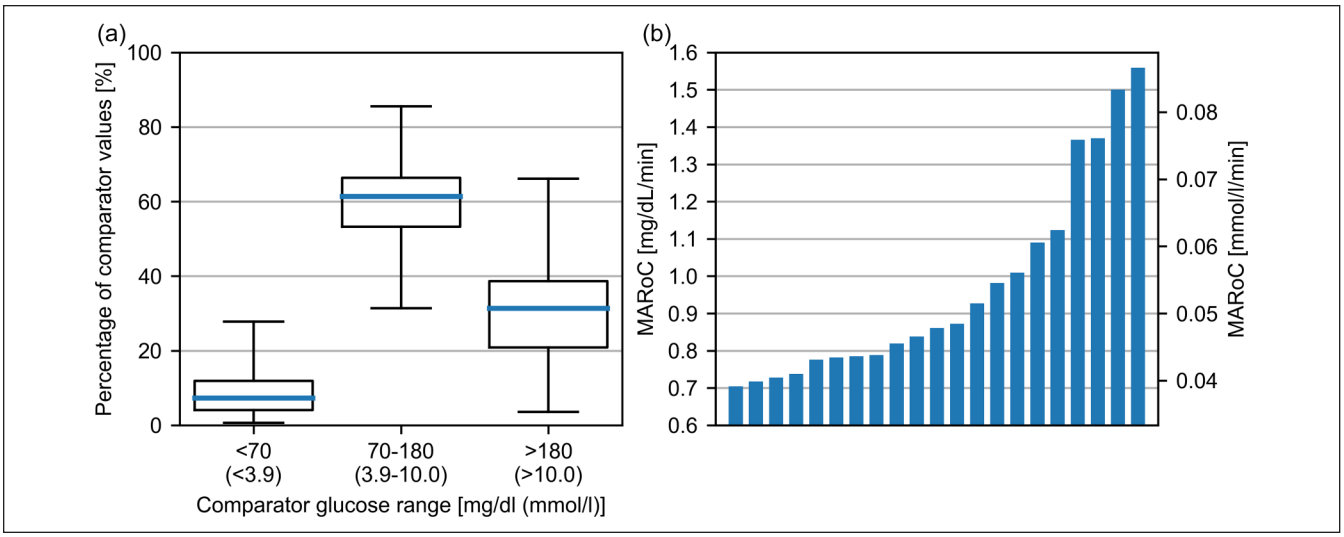

Figure 8. (a) Boxplots of percentages of comparator values in the TIR bins from 44 studies. The whiskers indicate minimum/maximum percentages. If multiple comparators were used, then only the distribution of venous measurements during in-clinic visits was analyzed. If the distribution was specified for each CGM system separately, the data were pooled. (b) Estimated mean absolute rate of change (MARoC) of comparator glucose concentrations from results reported in 21 studies.

In general, the level of reported details on the applied manipulation procedures varied greatly between studies. However, in the vast majority of the studies, the procedures were not described in sufficient detail to allow the independent development of a similar procedure. We thus encourage authors to provide sufficiently detailed descriptions of these procedures.

Data Analysis and Statistical Performance Evaluation

Comparator Data Characterization

Considering the previously described incomplete reporting of glucose manipulation procedures and the well-known impact of glucose concentrations and their dynamics on CGM accuracy, it is of utmost importance that the distribution of comparator data is sufficiently characterized. Such a characterization is the most suitable way to judge whether the performance results of studies, in particular accuracy, are comparable. Furthermore, this characterization should be considered the result of the general study design, comparator method selection, and testing procedures and should thus be reported and discussed independently from the endpoints of the study (Table 1).

There are two main facets of the comparator data characteristics: the distribution of glucose concentrations themselves and the distribution of glucose concentration RoCs. When it comes to the distribution of glucose concentrations, this review found that only 55.0% of the included studies provided a suitable description. This was defined as reporting the percentages of comparator data points in at least three different glucose ranges. Here it should be pointed out that several recent studies reported the distribution of CGM values instead of comparator values.[90,96,123,127,131,138,140,155,158,160]

However, this cannot be recommended as the CGM value distribution is dependent on CGM accuracy and thus impairs the comparison of testing procedures among studies.[168] The most common choice for the glucose ranges were the established time in range (TIR) bins <70 mg/dL (<3.9 mmol/L), 70 to 180 mg/dL (3.9-10.0 mmol/L), and >180 mg/dL (>10 mmol/L), used in 62.0% of studies that reported a distribution. The remaining studies used various bins, thus making comparisons of different studies very difficult. In particular, it was not sensible to systematically assess whether recommendations of the POCT05 of having >8% of comparator values <80 mg/dL (4.4 mmol/L) and >5% of comparator values >300 mg/dL (>16.7 mmol/L)[4] were followed.

An in-depth examination of the 44 studies reporting comparator values in the TIR bins found a great variability among studies with a median percentage of comparator values of 7.3% <70 mg/dL (<3.9 mmol/L), 61.4% between 70 and 180 mg/dL (3.9-10.0 mol/L), and 31.4% >180 mg/dL (>10 mmol/L). More details are displayed in Figure 8a.

The comparator glucose concentration RoCs have an equally important influence on the observed accuracy of a CGM system. However, a meaningful calculation of RoCs is only possible if comparator measurements were performed with a sufficiently high frequency. As mentioned before, this review chose a threshold of 20 minutes. Here, it was found that out of the 83 studies with a sampling interval of 20 minutes or less, only 21 (25.3%) reported a distribution of comparator RoCs, with some studies only reporting the distribution of CGM RoCs which cannot be recommended as explained before. Among the 21 studies that reported a comparator RoC distribution, the bins used were highly inconsistent, which

compromises the comparability of the results. To overcome this, the reported comparator RoC distributions were used to estimate the mean absolute rate of change (MARoC) in mg/dL/min (mmol/L/min) of comparator glucose concentrations (the methodology is described in the Supplemental Material). This is a single number that can characterize the extent of fast glucose changes, both positive and negative, in the studies. We therefore propose that future studies report this parameter alongside a suitable description of the comparator RoC distribution.

The results are depicted in Figure 8b and show that most MARoCs can be found between 0.7 and 1.0 mg/dL/min (0.039 and 0.056 mmol/L/min), with some studies reaching considerably higher MARoCs above 1.3 mg/dL/min (0.072 mg/dl/min).

When reporting the distribution of comparator glucose concentrations and their RoCs, it should be noted that these two quantities have a temporal association and it is quite difficult to visualize the actual glucose profiles that occurred during in-clinic sessions by just inspecting the distributions. In this context, we found that 37.6% of studies with in-clinic sessions reported the time course of glucose concentrations of at least one subject to illustrate the glycemic conditions during the in-clinic sessions.

Data Exclusion

There may be various reasons to exclude data from the evaluation, both recorded with the CGM systems and the comparators. These include deviations from the testing protocol and technical malfunctions. Furthermore, data exclusion can have a considerable impact on the resulting performance and should therefore be described in detail.

CGM-Comparator Pairing

Typically, CGM systems do not automatically store glucose concentrations every minute but use fixed recording intervals of 5 to 15 minutes, depending on the system used. Additionally, there are intermittently scanned CGM systems, for which measurements can be recorded at any time. This creates multiple approaches for pairing the recorded CGM values with comparator measurements, the choice of which has been shown to affect performance results.[117] We therefore recommend describing the method for pairing in sufficient detail, also because most studies (51.9%) did not provide such a description. If the pairing method was described, then the following options and their prevalence were observed (considering that more than one option could be used, depending on the CGM systems):

- Pairing the CGM value that was recorded closest in time, ie, either before or after, to the comparator value. This approach was the most common and made up 45.5% of all pairing methods used.
- Pairing the CGM value that was recorded simultaneously or after the comparator values. This approach was used in 27.3% of all pairing methods.
- Using linear interpolation to calculate and pair a CGM or comparator value that is concomitant with the recorded CGM or comparator value. This approach was used in 21.2% of all pairing methods.
- Pairing the CGM value that was recorded simultaneously or before the comparator values. This was the least common approach and was used in 6.1% of all pairing methods.

Statistical Accuracy Evaluation

The most important endpoint of a clinical CGM performance evaluation is accuracy, where it is common to distinguish between point accuracy, ie, the extent of agreement between CGM and comparator values at single points in time, and trend accuracy, ie, the ability of the CGM system to correctly indicate the direction and magnitude of glucose concentration changes. Furthermore, a distinction between analytical and clinical accuracy can be made (Table 1).

Numerous articles have already discussed the various parameters and approaches that can be used to characterize point accuracy.[5,12,168,169-171] Most of the approaches describe analytical accuracy, and we refer to those articles for detailed discussions of various approaches. This review identified the following nine approaches, the prevalence of which is summarized in Figure 9:

- Mean and/or median absolute relative difference (ARD) (94.6%) between CGM and comparator measurements. Despite some criticism,[5,172] the mean ARD (often abbreviated as MARD) remains one of the most frequently used parameters to assess point accuracy. Here it should be mentioned that mean and median ARDs have a different statistical interpretation and cannot be compared with each other. Studies where both mean and median ARDs have been reported show that median ARDs can be between 2.4[156] and 7[19] percentage points lower than mean ARDs. Another aspect of mean/median ARD calculation is whether all CGM-comparator data pairs are used at once (often referred to as aggregate mean/median ARD), or whether the mean/median ARDs of individual sensors are calculated first, and then a single number is calculated as mean/median of these sensor-specific mean/median ARDs. However, while the differences in mean/median results between these two approaches are typically small, the associated measure of variability (standard deviation or interquartile range) for the aggregate mean/median ARD is typically much larger. The evolution and in particular a trend toward a decrease of MARDs over time are shown in Figure 10. A table of

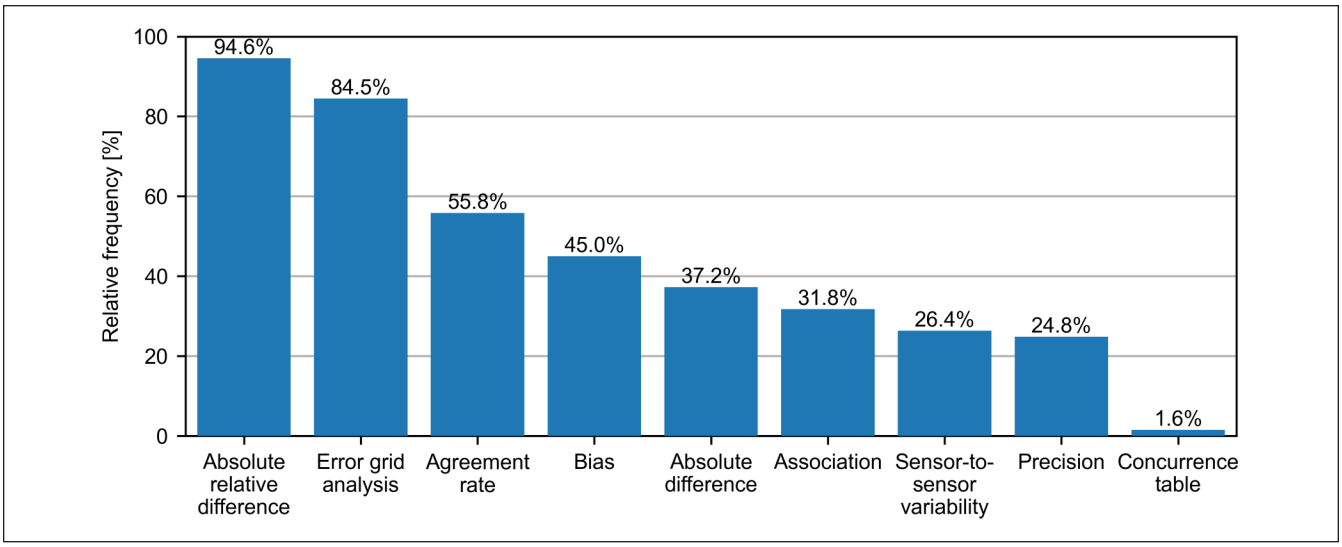

Figure 9. Parameters and methods to characterize point accuracy and their prevalence identified in the 129 studies included in this review.

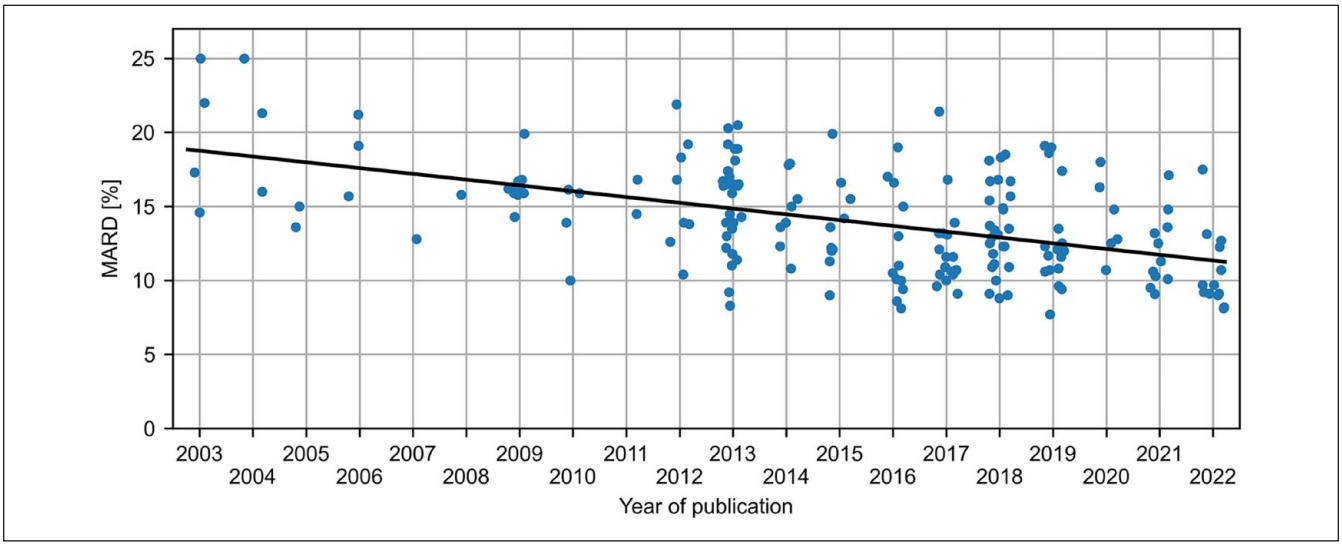

Figure 10. MARDs reported in the 129 studies included in this review with respect to their year of publication (n=189).
The data points were spread out within each year to better reflect the number of reported MARDs. Included were only MARD results reported over the full glucose range. If MARD values in specific studies were reported for different comparator compartments, calibration algorithms, or insertion sites, then they were included separately.

individual MARDs and the associated CGM systems is provided in the Supplemental Material.

- Error grid analysis (84.5%). It is the only established approach to characterize clinical accuracy, indicating the share of deviations between CGM and comparator values in predefined zones of clinical impact, ranging from none to severe. To this date, four different error grid analyses have been proposed and their usage over time is demonstrated in Figure 11a. Clarke et al,[173] consensus,[174] and surveillance[175] error grids focus on clinical point accuracy, whereas the continuous glucose error grid analysis (CG-EGA)[176] was specifically developed for CGM systems and incorporates both point and trend accuracy. An increase in the use of the consensus error grid can be observed after 2015, which is most likely due to its incorporation in the standard for BGMS published by the International Organization for Standardization (ISO).[162] Conversely, the use of the CG-EGA has declined over time. Here it should be mentioned that the extent to which

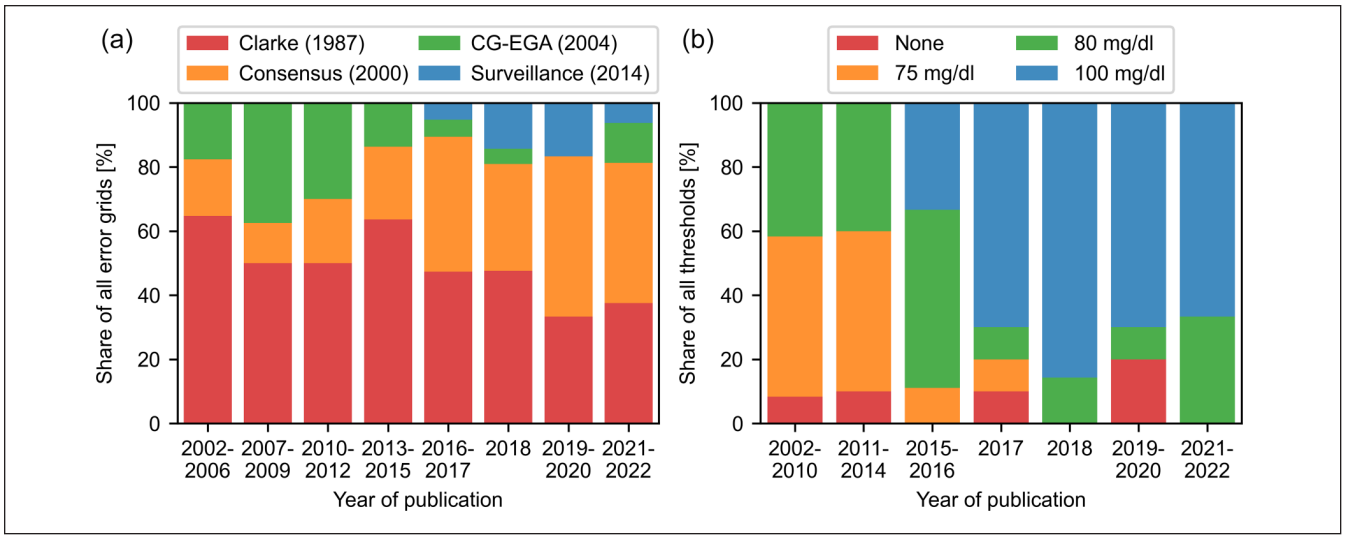

Figure 11. Usage of different error grids (a) and thresholds for agreement rates (b) expressed as share of all error grids/thresholds in specific time periods.

In both panels, the time intervals were chosen so that an approximately equal number of error grid/threshold usages occur in every time interval. In panel (a), the four types of error grids are the Clarke error grid, introduced by Clarke et al,[172] the consensus error grid, introduced by Parkes et al,[173] the CG-EGA, introduced by Kovatchev et al,[175] and the surveillance error grid, introduced by Klonoff et al.[174] In panel (b), the thresholds for switching between absolute (below threshold) and relative (above threshold) differences are characterized. "None" means that only relative differences were used.

CG-EGA results were reported varied greatly between articles. In general, the fact that four similar methods to characterize clinical accuracy are used to this day exemplifies the problem of comparing study results with each other.

- Agreement rates (55.8%). They indicate the share of CGM measurements within certain limits, eg, ±15 mg/dL (±0.83 mmol/L) or %, of comparator measurements. If a single agreement rate is calculated from all CGM-comparator pairs over the entire glucose range, then it is very common to switch from relative to absolute difference limits at certain thresholds of lower glucose concentrations to account for the glucose concentration dependency of CGM-comparator deviations. The use of these different thresholds has changed over time which is demonstrated in Figure 11b. The figure demonstrates that the use of 100 mg/dL (5.55 mmol/l) as a threshold has become dominant from 2015 onward which coincides with the update of the ISO 15197 standard for BGMS using this threshold.[162] The use of 80 mg/dL (4.44 mmol/L) is recommended in the POCT05 guideline,[4] and the FDA iCGM requirements only use relative differences ("none" in Figure 11b).[10]
- Mean and/or median bias (both absolute and/or relative) (45.0%) between CGM and comparator measurements, often combined with a Bland-Altman analysis (22.5%). These parameters can indicate whether CGM measurements systematically over- or underestimate comparator blood glucose concentrations, but bias alone can also mask a large amount of imprecision.
- Mean and/or median absolute difference (37.2%) between CGM and comparator measurements. These parameters are similar to mean/median ARDs and are commonly used to characterize the deviations in low glucose ranges.
- Analysis of association between CGM and comparator measurements using correlation and/or regression (31.8%). This approach was very common for early-generation CGM systems to demonstrate their general ability to provide information on glucose concentrations. However, as systems have improved, this analysis has mostly become obsolete.
- Variability in accuracy from sensor to sensor (26.4%). This is a very important aspect of CGM point accuracy and can be characterized, eg, by the distribution of individual sensor MARDs.
- Precision (24.8%) in the form of the mean or median paired ARD between two sensors and/or other approaches. Note that this parameter can only be calculated if at least two sensors from the same system are worn in parallel.
- Concurrence tables (1.6%). They divide the CGM and comparator data in various bins across the glucose range and indicate the number of CGM-comparator pairs lying within each possible combination of CGM and comparator bins. Although this approach was suggested in the first and second editions of the POCT05 guideline,[9,4] it was only reported in the manuscripts of two studies.[139,158]

Only 27 studies reported any trend accuracy results, the vast majority were published in 2015 or earlier. The rate error

grid, a part of the previously mentioned CG-EGA,[176] was most common and used to generate 51.9% of trend accuracy results. Two other approaches for trend accuracy assessment could be identified: the calculation of rate deviations (29.6%) analogous to deviations in the context of point accuracy and concurrence tables (18.5%), both of which are recommended in the POCT05 guideline.[4]

Another important aspect of the statistical accuracy evaluation is the reporting of confidence intervals of the previously mentioned parameters, which has gained relevance since the FDA iCGM requirements use confidence intervals to define their minimum acceptance criteria.[10] This review found that less than 25% of studies reported confidence intervals on any of the previously mentioned accuracy parameters. We recommend that all point estimates of accuracy parameters are reported with a suitably chosen and calculated confidence interval. For that, the specific properties of the data from CGM performance studies have to be taken into account, such as the previously mentioned interdependency of closely sampled data and clustering of data within CGM sensors.[177]

Other Aspects of CGM System Performance

Closely related to point accuracy is the stability of the sensors. In particular, the stability of accuracy over the sensor lifetime was reported in 38.8% of studies, often by stratifying one specific accuracy parameter, eg, MARD or agreement rate by the sensor wear day. The calibration stability, ie, an analysis of the accuracy with respect to the time after calibration, was only reported in two studies.[108,156]

Another aspect closely related to accuracy is the reliability of hypo- and hyperglycemic threshold alerts which has been reported in 24.8% of studies, respectively. Here, it was common to calculate the percentage of correct alarms also known as sensitivity (CGM values below/above a certain threshold concurrent with the comparator measurement below/above the same threshold) and correct detections also known as specificity (comparator measurement below/above a certain threshold concurrent with a CGM value below/above the same threshold), recommended by the POCT05 guideline. The reliability of predictive threshold alerts was assessed in 4.7% of studies. An analysis of the correlation between accuracy parameters and subject characteristics, such as body mass index, age, or duration of diabetes, was carried out in 10.1% of studies.

The technical reliability of the CGM system in terms of sensor survival was reported in 28.7% of studies. This can be characterized by several approaches, such as the percentage of sensors that reached the end of their lifetime or the survival probability calculated with the Kaplan-Meier method. The availability of data, ie, the ability of the CGM system to provide the expected number of glucose data without interruptions was reported in only 6.2% of studies. Both sensor survival and data availability were often accompanied by a description of device deficiencies that occurred during the study. Finally, the lag time between CGM and comparator data was reported in 17.8% of studies. However, a variety of different methods to determine the lag time was used.

To provide an indication of CGM system safety, any adverse events and adverse device effects that were related to the investigated CGM systems during the study were reported in 46.5% of studies. A related aspect is the satisfaction of the users with the CGM systems, which is typically assessed through user questionnaires inquiring on features such as wear comfort, pain of sensor insertion, and user-friendliness. This was reported in 24.0% of studies.

Finally, a particular strength of studies examining multiple CGM systems worn by the same subjects is the ability to compare various therapy parameters such as the TIR or glucose management indicator among CGM systems. Examples of this analysis can be found in the studies by Bonora et al,[95] Boscari et al,[154] and Yeoh et al,[161] concluding that, in particular, the time below range can be considerably different between CGM systems.

Recommendations for Reporting

A summary of our recommendations for reporting the previously discussed aspects of general study design, CGM system use, comparator measurement, testing procedures, and data analysis/statistical performance evaluation is provided in Table 2.

Limitations of This Review

A limitation of this review is that a more concise, quantitative analysis of the impact of various study design options on the observed performance was not carried out, because the number of studies examining the exact same CGM system was limited. Furthermore, the inclusion criteria of selected studies were chosen fairly broad, which might have contributed to the large observed variability. If the review had been limited to studies, eg, published within the last ten years or with a large subject population, more homogenous study designs would probably have been found.

Conclusion

This scoping review analyzed reports of 129 clinical studies published between 2002 and 2022 evaluating the performance of CGM systems. In particular, various aspects of study design were analyzed. We determined that, as a group, these studies are characterized by two important factors that interfere with attaining a clear idea of individual system performance as well as comparing the performance between systems: (1) the various options for protocol design have the potential to significantly impact study outcomes and cause inconsistency of study design and (2) crucial information

Table 2. Recommendations for Reporting the Design and Results of a Clinical CGM Performance Evaluation by the WG-CGM Established by the IFCC.

Recommendations for reporting a clinical CGM performance evaluation
General study design
• Ethical approval
• Inclusion/exclusion criteria including possible interfering medications
• Subject population characteristics
- Demographic (age, sex, BMI, HbA1c, and ethnicity)
- Diabetes type
- Diabetes treatment regime
• Sample size justification
• Number of investigational sites (mono- or multicenter)
• Role of funding source
CGM system use
• Number of systems
• Number of sensors used simultaneously
• Number of sensors used consecutively
• Number of planned sensor wear days and intended sensor lifetime
• Manufacturing lot(s) of sensors
• Source of CGM devices and sensors
• Protocol for insertion and possible replacement of sensors in case of failure
• Sensor insertion location, within or outside intended use
• Calibration protocol, including the source of calibrators used and concentration of the calibrators
Comparator measurement approach
• Sampling compartment (capillary, venous, and arterialized-venous)
• Details of arterialization protocol (if applicable)
• Glucose measurement method and device for comparator data
• Procedures for harmonization of comparator measurements (if applicable)
• Compliance with analytical performance specifications proposed by Pleus et al[163]
Testing procedures
• Study setting (in-clinic/free-living)
• Number of in-clinic sessions/days in relation to sensor lifetime
• Duration of in-clinic sessions
• Sampling schedule of comparator measurements
• Number of planned comparator measurements per subject and in total
• Detailed description of glucose manipulation protocol
Data analysis and statistical performance evaluation
• Comparator data characteristics[a]
- Glucose concentrations
- RoC
- Time course of comparator glucose concentrations
• Exclusion of data
• Pairing of CGM and comparator data
• Point accuracy[a]
• Trend accuracy[a]
• Threshold alert reliability[a]
• Predictive alert reliability[a]
• Technical reliability[a]
• Occurrence of adverse device effects[a]
• User satisfaction (if applicable)[a]

Abbreviations: CGM, continuous glucose monitoring; BMI, body mass index.
[a]The technical details on how to report these aspects lie beyond the scope of this review and will be the subject of a future publication.

about the study design is frequently not reported, jeopardizing the integrity and validity of the study findings.

The two goals of this review were first to assess the clarity with which clinical evaluations of CGM performance have been reported and second to stimulate interest in developing a future standard for reporting studies of CGM performance. Regarding the first goal, we believe that this analysis explains what information journal editors, reviewers, and regulators currently lack but need to intelligently assess the quality of a CGM performance evaluation. By advocating for adherence

to our reporting recommendations outlined in this article (Table 2), we hope to see all relevant information about the study design (including comparator data characteristics) reported in a transparent and consistent manner. Regarding the second goal, we believe that this review can encourage the development of a standard for designing CGM performance studies to improve the clarity of their results. The result of both goals will be a greater understanding of the performance of CGMs, which are becoming an increasingly indispensable tool for managing diabetes.

Abbreviations

ARD, absolute relative difference; BGMS, blood glucose monitoring system; CG-EGA, continuous glucose error grid analysis; CGM, continuous glucose monitoring; FDA, U.S. Food and Drug Administration; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine; ISF, interstitial fluid; ISO, International Organization of Standardization; MARoC, mean absolute rate of change; RoC, rate of change; TIR, time in range; YSI, Yellow Springs Instruments.

Acknowledgments

The authors would like to thank the Diabetes Center Berne and the International Federation of Clinical Chemistry and Laboratory Medicine for supporting this work.

Declaration of Conflicting Interests

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 the general manager and medical director of the Institute for Diabetes Technology (Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft mbH an der Universität Ulm, Ulm, Germany), which carries out clinical studies, eg, with medical devices for diabetes therapy on its own initiative and on behalf of various companies. G.F./IfDT have received research support, speakers' honoraria, or consulting fees in the last three years from Abbott, Ascensia, Berlin Chemie, Boydsense, Dexcom, Lilly, Metronom, Medtronic, Menarini, MySugr, Novo Nordisk, PharmaSens, Roche, Sanofi, Terumo. M.E., D.W., S.P., S.W., and C.H. are employees of IfDT. R.S. is the chair of the Clinical Chemistry Department of Isala which carries out clinical studies, eg, with medical devices for diabetes therapy on its own initiative and on behalf of various companies. R.S. has received speakers' honoraria or consulting fees in the last three years from Roche and Menarini. J.J. has for the last three years been a lecturer/member of the scientific advisory board for Abbott, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Medtronic, Nordic InfuCare, NovoNordisk A/S, and Sanofi. R.H. is an employee of Roche Diabetes Care GmbH. L.W. has received funding from the Diabetes Center Berne. E.E.B. and K.M. have no disclosures. D.C.K. is a consultant for Atropos Health, Better Therapeutics, Eoflow, Integrity, Lifecare, Nevro, Novo, Sanofi, and Thirdwayv. N.T. is a consultant for Roche Diagnostics and Radiometer. J.H.N. has received research support from Abbott. P.D. is a board member of PharmaSens. A.T. is a freelance consultant. He has received fees for lectures or consultancy fees from Abbott, Berlin Chemie, Dexcom, Evivamed, Menarini, Novo Nordisk, Roche, and Sanofi in the last three years.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This literature review and the associated scientific writing activities were funded by the Diabetes Center Berne.

ORCID iDs

Guido Freckmann https://orcid.org/0000-0002-0406-9529
Manuel Eichenlaub https://orcid.org/0000-0003-2150-3160
Delia Waldenmaier https://orcid.org/0000-0003-3280-2369
Stefan Pleus https://orcid.org/0000-0003-4629-7754
Lilian Witthauer https://orcid.org/0000-0001-9459-875X
Johan Jendle https://orcid.org/0000-0003-1025-1682
Rolf Hinzmann https://orcid.org/0000-0002-8419-4740
Andreas Thomas https://orcid.org/0000-0002-6549-2793
Konstantinos Makris https://orcid.org/0000-0002-7896-9028
Nam Tran https://orcid.org/0000-0003-1565-0025
David C. Klonoff https://orcid.org/0000-0001-6394-6862
James H. Nichols https://orcid.org/0000-0002-3652-1612

Supplemental Material

Supplemental material for this article is available online.

References

1. Elbalshy M, Haszard J, Smith H, et al. Effect of divergent continuous glucose monitoring technologies on glycaemic control in type 1 diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. *Diabet Med*. 2022;39(8):e14854.
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. American Diabetes Association. 7. Diabetes technology: standards of medical care in diabetes—2021. *Diabetes Care*. 2021;44(supplement 1):S85.
4. Clinical and Laboratory Standards Institute (CLSI). *Performance Metrics for Continuous Interstitial Glucose Monitoring* (CLSI Guideline POCT05). 2nd ed. Wayne, PA: CLSI; 2020
5. 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.
6. Bailey TS, Alva S. Landscape of continuous glucose monitoring (CGM) and integrated CGM: accuracy considerations. *Diabetes Technol Ther*. 2021;23(S3):S5-S11.
7. Freckmann G, Nichols JH, Hinzmann R, et al. Standardization process of continuous glucose monitoring: traceability and performance. *Clin Chim Acta*. 2021;515:5-12.
8. Nichols JH, Klonoff DC. The need for performance standards for continuous glucose monitors. *J Diabetes Sci Technol*. 2007;1(1):92-94.

9. Clinical and Laboratory Standards Institute (CLSI). *Performance Metrics for Continuous Interstitial Glucose Monitoring; Approved Guideline* (CLSI document POCT05-A). Wayne, PA: CLSI; 2008
10. U.S.Food and Drug Administration. *Evaluation of Automatic Class III Designation for Dexcom G6 Continuous Glucose Monitoring System: Decision Summary*. https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170088.pdf. Published online 2018. Accessed May 24, 2022.
11. 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.
12. Pemberton JS, Wilmot EG, Barnard-Kelly K, et al. CGM accuracy: contrasting CE marking with the governmental controls of the USA (FDA) and Australia (TGA)—a narrative review. *Diabetes Obes Metab*. 2023;24:916-939.
13. Welsh JB, Psavko S, Zhang X, Gao P, Balo AK. Comparisons of fifth-, sixth-, and seventh-generation continuous glucose monitoring systems [published online ahead of print June 13, 2022]. *J Diabetes Sci Technol*. doi: 10.1177/19322968221099879.
14. Muñoz Fabra E, Díez JL, Bondia J, Laguna Sanz AJ. A comprehensive review of continuous glucose monitoring accuracy during exercise periods. *Sensors (Basel)*. 2021;21(2):479.
15. Galindo RJ, Umpierrez GE, Rushakoff RJ, et al. Continuous glucose monitors and automated insulin dosing systems in the hospital consensus guideline. *J Diabetes Sci Technol*. 2020;14(6):1035-1064.
16. Thiese MS. Observational and interventional study design types: an overview. *Biochem Med (Zagreb)*. 2014;24(2): 199-210.
17. International Federation of Clinical Chemistry and Laboratory Medicine Scientific Division Working Group on Continuous Glucose Monitoring (WG-CGM). https://ifcc.org/ifcc-scientific-division/sd-working-groups/wg-cgm/.
18. Maran A, Crepaldi C, Tiengo A, et al. Continuous subcutaneous glucose monitoring in diabetic patients: a multicenter analysis. *Diabetes Care*. 2002;25(2):347-352.
19. The Diabetes Research in Children Network (DIRECNET) Study Group. The accuracy of the CGMS in children with type 1 diabetes: results of the diabetes research in children network (DirecNet) accuracy study. *Diabetes Technol Ther*. 2003;5(5):781-789.
20. The Diabetes Research in Children Network (DIRECNET) Study Group. The accuracy of the GlucoWatch G2 biographer in children with type 1 diabetes: results of the diabetes research in children network (DirecNet) accuracy study. *Diabetes Technol Ther*. 2003;5(5):791-800.
21. The Diabetes Research in Children Network (DIRECNET) Study Group. Accuracy of the GlucoWatch G2 biographer and the continuous glucose monitoring system during hypoglycemia. *Diabetes Care*. 2004;27(3):722-726.
22. The Diabetes Research in Children Network (DIRECNET) Study Group. Accuracy of the modified continuous glucose monitoring system (CGMS) sensor in an outpatient setting: results from a diabetes research in children network (DirecNet) study. *Diabetes Technol Ther*. 2005;7(1):109-114.
23. Djakouré-Platonoff C, Radermercker R, Reach G, Slama G, Selam JI. Accuracy of the continuous glucose monitoring system in inpatient and outpatient conditions. *Diabetes Metab*. 2003;29(2 pt 1):159-162.
24. Feldman B, Brazg R, Schwartz S, Weinstein R. A continuous glucose sensor based on wired enzyme technology: results from a 3-day trial in patients with type 1 diabetes. *Diabetes Technol Ther*. 2003;5(5):769-779.
25. 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.
26. Guerci B, Floriot M, Böhme P, et al. Clinical performance of CGMS in type 1 diabetic patients treated by continuous subcutaneous insulin infusion using insulin analogs. *Diabetes Care*. 2003;26(3):582-589.
27. Sachedina N, Pickup JC. Performance assessment of the Medtronic-Minimed continuous glucose monitoring system and its use for measurement of glycaemic control in type 1 diabetic subjects. *Diabet Med*. 2003;20(12):1012-1015.
28. Bode B, Gross K, Rikalo N, et al. Alarms based on real-time sensor glucose values alert patients to hypo- and hyperglycemia: the guardian continuous monitoring system. *Diabetes Technol Ther*. 2004;6(2):105-113.
29. Garg SK, Schwartz S, Edelman SV. Improved glucose excursions using an implantable real-time continuous glucose sensor in adults with type 1 diabetes. *Diabetes Care*. 2004;27(3):734-738.
30. Nybäck-Nakell A, von Heijne M, Adamson U, Lins PE, Landstedt-Hallin L. Accuracy of continuous nocturnal glucose screening after 48 and 72 hours in type 2 diabetes patients on combined oral and insulin therapy. *Diabetes Metab*. 2004;30(6):517-521.
31. Clarke WL, Anderson S, Farhy L, et al. Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose-error grid analysis. *Diabetes Care*. 2005;28(10):2412-2417.
32. Wentholt IM, Vollebregt MA, Hart AA, Hoekstra JB, DeVries JH. Comparison of a needle-type and a microdialysis continuous glucose monitor in type 1 diabetic patients. *Diabetes Care*. 2005;28(12):2871-2876.
33. Buckingham BA, Kollman C, Beck R, et al; Diabetes Research in Children Network (Direcnet) Study Group. Evaluation of factors affecting CGMS calibration. *Diabetes Technol Ther*. 2006;8(3):318-325.
34. Garg S, Zisser H, Schwartz S, et al. Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor: a randomized controlled trial. *Diabetes Care*. 2006;29(1):44-50.
35. Garg S, Jovanovic L. Relationship of fasting and hourly blood glucose levels to HbA1c values: safety, accuracy, and improvements in glucose profiles obtained using a 7-day continuous glucose sensor. *Diabetes Care*. 2006;29(12): 2644-2649.
36. Kubiak T, Wörle B, Kuhr B, et al. Microdialysis-based 48-hour continuous glucose monitoring with GlucoDay: clinical performance and patients' acceptance. *Diabetes Technol Ther*. 2006;8(5):570-575.

37. Wilhelm B, Forst S, Weber MM, Larbig M, Pfützner A, Forst T. Evaluation of CGMS during rapid blood glucose changes in patients with type 1 diabetes. *Diabetes Technol Ther*. 2006;8(2):146-155.
38. Wilson DM, Beck RW, Tamborlane WV, et al. The accuracy of the FreeStyle Navigator continuous glucose monitoring system in children with type 1 diabetes. *Diabetes Care*. 2007;30(1):59-64.
39. Weinstein RL, Schwartz SL, Brazg RL, Bugler JR, Peyser TA, McGarraugh GV. Accuracy of the 5-day FreeStyle Navigator Continuous Glucose Monitoring System: comparison with frequent laboratory reference measurements. *Diabetes Care*. 2007;30(5):1125-1130.
40. McGarraugh GV, Clarke WL, Kovatchev BP. Comparison of the clinical information provided by the FreeStyle Navigator continuous interstitial glucose monitor versus traditional blood glucose readings. *Diabetes Technol Ther*. 2010;12(5):365-371.
41. Bode B, Silver M, Weiss R, Martin K. Evaluation of a continuous glucose monitoring system for home-use conditions. *Manag Care*. 2008;17(8):40-45.
42. The Diabetes Research in Children Network (DIRECNET) Study Group . The accuracy of the Guardian (R) RT continuous glucose monitor in children with type 1 diabetes. *Diabetes Technol Ther*. 2008;10(4):266-272.
43. Calhoun P, Lum J, Beck RW, Kollman C. Performance comparison of the medtronic sof-sensor and enlite glucose sensors in inpatient studies of individuals with type 1 diabetes. *Diabetes Technol Ther*. 2013;15(9):758-761.
44. Kovatchev B, Anderson S, Heinemann L, Clarke W. Comparison of the numerical and clinical accuracy of four continuous glucose monitors. *Diabetes Care*. 2008;31(6): 1160-1164.
45. Mastrototaro J, Shin J, Marcus A, Sulur G; STAR 1 Clinical Trial Investigators. The accuracy and efficacy of real-time continuous glucose monitoring sensor in patients with type 1 diabetes. *Diabetes Technol Ther*. 2008;10(5):385-390.
46. Bailey T, Zisser H, Chang A. New features and performance of a next-generation SEVEN-day continuous glucose monitoring system with short lag time. *Diabetes Technol Ther*. 2009;11(12):749-755.
47. Garg SK, Smith J, Beatson C, Lopez-Baca B, Voelmle M, Gottlieb PA. Comparison of accuracy and safety of the SEVEN and the Navigator continuous glucose monitoring systems. *Diabetes Technol Ther*. 2009;11(2):65-72.
48. Kamath A, Mahalingam A, Brauker J. Analysis of time lags and other sources of error of the DexCom SEVEN continuous glucose monitor. *Diabetes Technol Ther*. 2009;11(11): 689-695.
49. Mazze RS, Strock E, Borgman S, Wesley D, Stout P, Racchini J. Evaluating the accuracy, reliability, and clinical applicability of continuous glucose monitoring (CGM): Is CGM ready for real time. *Diabetes Technol Ther*. 2009;11(1):11-18.
50. Nielsen JK, Christiansen JS, Kristensen JS, et al. Clinical evaluation of a transcutaneous interrogated fluorescence lifetime-based microsensor for continuous glucose reading. *J Diabetes Sci Technol*. 2009;3(1):98-109.
51. Nielsen JK, Freckmann G, Kapitza C, et al. Glucose monitoring by microdialysis: performance in a multicentre study. *Diabet Med*. 2009;26(7):714-721.
52. Zisser HC, Bailey TS, Schwartz S, Ratner RE, Wise J. Accuracy of the SEVEN continuous glucose monitoring system: comparison with frequently sampled venous glucose measurements. *J Diabetes Sci Technol*. 2009;3(5):1146-1154.
53. Keenan DB, Cartaya R, Mastrototaro JJ. Accuracy of a new real-time continuous glucose monitoring algorithm. *J Diabetes Sci Technol*. 2010;4(1):111-118.
54. Meschi F, Bonfanti R, Rigamonti A, et al. Patients' evaluation of nocturnal hypoglycaemia with GlucoDay continuous glucose monitoring in paediatric patients. *Acta Diabetol*. 2010;47(4):295-300.
55. Valgimigli F, Lucarelli F, Scuffi C, Morandi S, Sposato I. Evaluating the clinical accuracy of GlucoMen(R)Day: a novel microdialysis-based continuous glucose monitor. *J Diabetes Sci Technol*. 2010;4(5):1182-1192.
56. McGarraugh G, Brazg R, Weinstein R. FreeStyle navigator continuous glucose monitoring system with TRUstart algorithm, a 1-hour warm-up time. *J Diabetes Sci Technol*. 2011;5(1):99-106.
57. Morrow L, Hompesch M, Tideman AM, et al. Evaluation of a novel continuous glucose measurement device in patients with diabetes mellitus across the glycemic range. *J Diabetes Sci Technol*. 2011;5(4):853-859.
58. Peoples T, Bailey T, Ronald B, et al. Accuracy performance of the Medtronic NexSensorTM for 6 days in an inpatient setting using abdomen and buttocks insertion sites. *J Diabetes Sci Technol*. 2011;5(2):358-364.
59. Francescato MP, Geat M, Stel G, Cauci S. Accuracy of a portable glucose meter and of a continuous glucose monitoring device used at home by patients with type 1 diabetes. *Clin Chim Acta*. 2012;413(1-2):312-318.
60. Keenan DB, Mastrototaro JJ, Zisser H, et al. Accuracy of the Enlite 6-day glucose sensor with guardian and Veo calibration algorithms. *Diabetes Technol Ther*. 2012;14(3): 225-231.
61. Luijf YM, Avogaro A, Benesch C, et al. Continuous glucose monitoring accuracy results vary between assessment at home and assessment at the clinical research center. *J Diabetes Sci Technol*. 2012;6(5):1103-1106.
62. Weinstock RS, Bristol S, Armenia A, Gesswein AC, Bequette BW, Willis JP. Pilot study of a prototype minimally invasive intradermal continuous glucose monitor. *J Diabetes Sci Technol*. 2012;6(6):1454-1463.
63. Zhou J, Lv X, Mu Y, et al. The accuracy and efficacy of real-time continuous glucose monitoring sensor in Chinese diabetes patients: a multicenter study. *Diabetes Technol Ther*. 2012;14(8):710-718.
64. Zueger T, Diem P, Mougiakakou S, Stettler C. Influence of time point of calibration on accuracy of continuous glucose monitoring in individuals with type 1 diabetes. *Diabetes Technol Ther*. 2012;14(7):583-588.
65. Bay C, Kristensen PL, Pedersen-Bjergaard U, Tarnow L, Thorsteinsson B. Nocturnal continuous glucose monitoring: accuracy and reliability of hypoglycemia detection in patients with type 1 diabetes at high risk of severe hypoglycemia. *Diabetes Technol Ther*. 2013;15(5):371-377.
66. Christiansen M, Bailey T, Watkins E, et al. A new-generation continuous glucose monitoring system: improved accuracy and reliability compared with a previous-generation system. *Diabetes Technol Ther*. 2013;15:881-888.

67. Garcia A, Rack-Gomer AL, Bhavaraju NC, et al. Dexcom G4AP: an advanced continuous glucose monitor for the artificial pancreas. *J Diabetes Sci Technol*. 2013;7(6): 1436-1445.
68. Nakamura K, Balo A. The accuracy and efficacy of the Dexcom G4 platinum continuous glucose monitoring system. *J Diabetes Sci Technol*. 2015;9(5):1021-1026.
69. Damiano ER, El-Khatib FH, Zheng H, Nathan DM, Russell SJ. A comparative effectiveness analysis of three continuous glucose monitors. *Diabetes Care*. 2013;36(2):251-259.
70. Freckmann G, Pleus S, Link M, Zschornack E, Klotzer HM, Haug C. Performance evaluation of three continuous glucose monitoring systems: comparison of six sensors per subject in parallel. *J Diabetes Sci Technol*. 2013;7(4):842-853.
71. Hoss U, Budiman ES, Liu H, Christiansen MP. Continuous glucose monitoring in the subcutaneous tissue over a 14-day sensor wear period. *J Diabetes Sci Technol*. 2013;7(5): 1210-1219.
72. Leelarathna L, Nodale M, Allen JM, et al. Evaluating the accuracy and large inaccuracy of two continuous glucose monitoring systems. *Diabetes Technol Ther*. 2013;15(2):143-149.
73. Luijf YM, Mader JK, Doll W, et al. Accuracy and reliability of continuous glucose monitoring systems: a head-to-head comparison. *Diabetes Technol Ther*. 2013;15(8):722-727.
74. Müller AJ, Knuth M, Nikolaus KS, Krivánek R, Küster F, Hasslacher C. First clinical evaluation of a new percutaneous optical fiber glucose sensor for continuous glucose monitoring in diabetes. *J Diabetes Sci Technol*. 2013;7(1):13-23.
75. O'Neal DN, Adhya S, Jenkins A, Ward G, Welsh JB, Voskanyan G. Feasibility of adjacent insulin infusion and continuous glucose monitoring via the Medtronic Combo-Set. *J Diabetes Sci Technol*. 2013;7(2):381-388.
76. Zschornack E, Schmid C, Pleus S, et al. Evaluation of the performance of a novel system for continuous glucose monitoring. *J Diabetes Sci Technol*. 2013;7(4):815-823.
77. Pleus S, Schmid C, Link M, et al. Performance evaluation of a continuous glucose monitoring system under conditions similar to daily life. *J Diabetes Sci Technol*. 2013;7(4):833-841.
78. Pleus S, Schoemaker M, Morgenstern K, et al. Rate-of-change dependence of the performance of two CGM systems during induced glucose swings. *J Diabetes Sci Technol*. 2015;9(4):801-807.
79. Zijlstra E, Heise T, Nosek L, Heinemann L, Heckermann S. Continuous glucose monitoring: quality of hypoglycaemia detection. *Diabetes Obes Metab*. 2013;15(2):130-135.
80. Bailey TS, Ahmann A, Brazg R, et al. Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. *Diabetes Technol Ther*. 2014;16(5):277-283.
81. Damiano ER, McKeon K, El-Khatib FH, Zheng H, Nathan DM, Russell SJ. A comparative effectiveness analysis of three continuous glucose monitors: the Navigator, G4 Platinum, and Enlite. *J Diabetes Sci Technol*. 2014;8(4):699-708.
82. Sinha M, McKeon KM, Parker S, et al. A comparison of time delay in three continuous glucose monitors for adolescents and adults. *J Diabetes Sci Technol*. 2017;11(6):1132-1137.
83. Jina A, Tierney MJ, Tamada JA, et al. Design, development, and evaluation of a novel microneedle array-based continuous glucose monitor. *J Diabetes Sci Technol*. 2014;8(3):483-487.
84. Mahmoudi Z, Jensen MH, Dencker Johansen M, et al. Accuracy evaluation of a new real-time continuous glucose monitoring algorithm in hypoglycemia. *Diabetes Technol Ther*. 2014;16(10):667-678.
85. Matuleviciene V, Joseph JI, Andelin M, et al. A clinical trial of the accuracy and treatment experience of the Dexcom G4 sensor (Dexcom G4 system) and Enlite sensor (guardian REAL-time system) tested simultaneously in ambulatory patients with type 1 diabetes. *Diabetes Technol Ther*. 2014;16(11):759-767.
86. Andelin M, Kropff J, Matuleviciene V, et al. Assessing the accuracy of continuous glucose monitoring (CGM) calibrated with capillary values using capillary or venous glucose levels as a reference. *J Diabetes Sci Technol*. 2016;10(4):876-884.
87. Mortellaro M, DeHennis A. Performance characterization of an abiotic and fluorescent-based continuous glucose monitoring system in patients with type 1 diabetes. *Biosens Bioelectron*. 2014;61:227-231.
88. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. *J Diabetes Sci Technol*. 2015;9(2):209-214.
89. Peyser TA, Nakamura K, Price D, Bohnett LC, Hirsch IB, Balo A. Hypoglycemic accuracy and improved low glucose alerts of the latest Dexcom G4 platinum continuous glucose monitoring system. *Diabetes Technol Ther*. 2015;17:548-554.
90. Welsh JB, Gao P, Derdzinski M, et al. Accuracy, utilization, and effectiveness comparisons of different continuous glucose monitoring systems. *Diabetes Technol Ther*. 2019;21(3):128-132.
91. 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.
92. Kropff J, Bruttomesso D, Doll W, et al. Accuracy of two continuous glucose monitoring systems: a head-to-head comparison under clinical research centre and daily life conditions. *Diabetes Obes Metab*. 2015;17(4):343-349.
93. Nørgaard K, Shin J, Welsh JB, Gjessing H. Performance and acceptability of a combined device for insulin infusion and glucose sensing in the home setting. *J Diabetes Sci Technol*. 2015;9(2):215-220.
94. Thabit H, Leelarathna L, Wilinska ME, et al. Accuracy of continuous glucose monitoring during three closed-loop home studies under free-living conditions. *Diabetes Technol Ther*. 2015;17(11):801-807.
95. Bonora B, Maran A, Ciciliot S, Avogaro A, Fadini GP. Head-to-head comparison between flash and continuous glucose monitoring systems in outpatients with type 1 diabetes. *J Endocrinol Invest*. 2016;39(12):1391-1399.
96. Laffel L. Improved accuracy of continuous glucose monitoring systems in pediatric patients with diabetes mellitus: results from two studies. *Diabetes Technol Ther*. 2016;18(suppl 2):S223-S233
97. McAuley SA, Dang TT, Horsburgh JC, et al. Feasibility of an orthogonal redundant sensor incorporating optical plus redundant electrochemical glucose sensing. *J Diabetes Sci Technol*. 2016;10(3):679-688.
98. Sharifi A, Varsavsky A, Ulloa J, et al. Redundancy in glucose sensing: enhanced accuracy and reliability of an

electrochemical redundant sensor for continuous glucose monitoring. *J Diabetes Sci Technol*. 2016;10(3):669-678.
99. Taleb N, Emami A, Suppere C, et al. Comparison of two continuous glucose monitoring systems, Dexcom G4 platinum and medtronic paradigm Veo Enlite system, at rest and during exercise. *Diabetes Technol Ther*. 2016;18(9):561-567.
100. Aberer F, Hajnsek M, Rumpler M, et al. Evaluation of subcutaneous glucose monitoring systems under routine environmental conditions in patients with type 1 diabetes. *Diabetes Obes Metab*. 2017;19(7):1051-1055.
101. Biagi L, Hirata Bertachi A, Conget I, et al. Extensive assessment of blood glucose monitoring during postprandial period and its impact on closed-loop performance. *J Diabetes Sci Technol*. 2017;11(6):1089-1095.
102. Christiansen MP, Garg SK, Brazg R, et al. Accuracy of a fourth-generation subcutaneous continuous glucose sensor. *Diabetes Technol Ther*. 2017;19(8):446-456.
103. Edge J, Acerini C, Campbell F, et al. An alternative sensor-based method for glucose monitoring in children and young people with diabetes. *Arch Dis Child*. 2017;102(6): 543-549.
104. Faccioli S, Del Favero S, Visentin R, et al. Accuracy of a CGM sensor in pediatric subjects with type 1 diabetes. *J Diabetes Sci Technol*. 2017;11(6):1147-1154.
105. Fokkert MJ, van Dijk PR, Edens MA, et al. Performance of the FreeStyle Libre Flash glucose monitoring system in patients with type 1 and 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2017;5(1):e000320.
106. Hughes J, Welsh JB, Bhavaraju NC, Vanslyke SJ, Balo AK. Stability, accuracy, and risk assessment of a novel subcutaneous glucose sensor. *Diabetes Technol Ther*. 2017;19(S3):S21-S24.
107. Ji L, Guo X, Guo L, Ren Q, Yu N, Zhang J. A multicenter evaluation of the performance and usability of a novel glucose monitoring system in Chinese adults with diabetes. *J Diabetes Sci Technol*. 2017;11(2):290-295.
108. Kropff J, Choudhary P, Neupane S, et al. Accuracy and longevity of an implantable continuous glucose sensor in the PRECISE study: a 180-day, prospective, multicenter, pivotal trial. *Diabetes Care*. 2017;40(1):63-68.
109. Kropff J, van Steen SC, deGraaff P, Chan MW, van Amstel RBE, DeVries JH. Venous, arterialized-venous, or capillary glucose reference measurements for the accuracy assessment of a continuous glucose monitoring system. *Diabetes Technol Ther*. 2017;19(11):609-617.
110. Kuroda A, Taniguchi S, Akehi Y, et al. Accuracy and time delay of glucose measurements of continuous glucose monitoring and bedside artificial pancreas during hyperglycemic and euglycemic hyperinsulinemic glucose clamp study. *J Diabetes Sci Technol*. 2017;11(6):1096-1100.
111. Ólafsdóttir AF, Attvall S, Sandgren U, et al. A clinical trial of the accuracy and treatment experience of the flash glucose monitor FreeStyle Libre in adults with type 1 diabetes. *Diabetes Technol Ther*. 2017;19(3):164-172.
112. Rumpler M, Mader JK, Fischer JP, et al. First application of a transcutaneous optical single-port glucose monitoring device in patients with type 1 diabetes mellitus. *Biosens Bioelectron*. 2017;88:240-248.
113. Biagi L, Bertachi A, Quirós C, et al. Accuracy of continuous glucose monitoring before, during, and after aerobic and anaerobic exercise in patients with type 1 diabetes mellitus. *Biosensors (Basel)*. 2018;8(1):22.
114. Boscari F, Galasso S, Facchinetti A, et al. FreeStyle Libre and Dexcom G4 Platinum sensors: accuracy comparisons during two weeks of home use and use during experimentally induced glucose excursions. *Nutr Metab Cardiovasc Dis*. 2018;28(2):180-186.
115. Charleer S, Mathieu C, Nobels F, Gillard P. Accuracy and precision of flash glucose monitoring sensors inserted into the abdomen and upper thigh compared with the upper arm. *Diabetes Obes Metab*. 2018;20(6):1503-1507.
116. Christiansen MP, Klaff LJ, Brazg R, et al. A prospective multicenter evaluation of the accuracy of a novel implanted continuous glucose sensor: PRECISE II. *Diabetes Technol Ther*. 2018;20:197-206.
117. Freckmann G, Link M, Pleus S, Westhoff A, Kamecke U, Haug C. Measurement performance of two continuous tissue glucose monitoring systems intended for replacement of blood glucose monitoring. *Diabetes Technol Ther*. 2018;20(8):541-549.
118. Moser O, Sternad C, Eckstein ML, et al. Performance of intermittently scanned continuous glucose monitoring systems in people with type 1 diabetes: a pooled analysis. *Diabetes Obes Metab*. 2022;522-529.
119. Giani E, Macedoni M, Barilli A, et al. Performance of the flash glucose monitoring system during exercise in youth with type 1 diabetes. *Diabetes Res Clin Pract*. 2018;146:321-329.
120. Hansen EA, Klee P, Dirlewanger M, et al. Accuracy, satisfaction and usability of a flash glucose monitoring system among children and adolescents with type 1 diabetes attending a summer camp. *Pediatr Diabetes*. 2018;19(7):1276-1284.
121. Massa GG, Gys I, Op't Eyndt A, et al. Evaluation of the FreeStyle® Libre Flash glucose monitoring system in children and adolescents with type 1 diabetes. *Horm Res Paediatr*. 2018;89(3):189-199.
122. Piona C, Dovc K, Mutlu GY, et al. Non-adjunctive flash glucose monitoring system use during summer-camp in children with type 1 diabetes: the free-summer study. *Pediatr Diabetes*. 2018;19(7):1285-1293.
123. 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.
124. Slover RH, Tryggestad JB, DiMeglio LA, et al. Accuracy of a fourth-generation continuous glucose monitoring system in children and adolescents with type 1 diabetes. *Diabetes Technol Ther*. 2018;20(9):576-584.
125. Staal OM, Hansen HMU, Christiansen SC, Fougner AL, Carlsen SM, Stavdahl Ø. Differences between flash glucose monitor and fingerprick measurements. *Biosensors (Basel)*. 2018;8(4):93.
126. Szadkowska A, Gawrecki A, Michalak A, Zozulinska-Ziolkiewicz D, Fendler W, Mlynarski W. Flash glucose measurements in children with type 1 diabetes in real-life settings: to trust or not to trust? *Diabetes Technol Ther*. 2018;20(1):17-24.
127. Wadwa RP, Laffel LM, Shah VN, Garg SK. Accuracy of a factory-calibrated, real-time continuous glucose monitoring system during 10 days of use in youth and adults with diabetes. *Diabetes Technol Ther*. 2018;20(6):395-402.

128. Zhou J, Zhang S, Li L, et al. Performance of a new real-time continuous glucose monitoring system: a multicenter pilot study. *J Diabetes Investig*. 2018;9(2):286-293.
129. Aronson R, Abitbol A, Tweden KS. First assessment of the performance of an implantable continuous glucose monitoring system through 180 days in a primarily adolescent population with type 1 diabetes. *Diabetes Obes Metab*. 2019;21(7): 1689-1694.
130. Cao B, Wang R, Gong C, et al. An evaluation of the accuracy of a flash glucose monitoring system in children with diabetes in comparison with venous blood blucose. *J Diabetes Res*. 2019;2019:4845729.
131. Christiansen MP, Klaff LJ, Bailey TS, Brazg R, Carlson G, Tweden KS. A prospective multicenter evaluation of the accuracy and safety of an implanted continuous glucose sensor: the PRECISION study. *Diabetes Technol Ther*. 2019;21:231-237.
132. Freckmann G, Link M, Kamecke U, Haug C, Baumgartner B, Weitgasser R. Performance and usability of three systems for continuous glucose monitoring in direct comparison. *J Diabetes Sci Technol*. 2019;13(5):890-898.
133. Kumagai R, Muramatsu A, Fujii M, et al. Comparison of glucose monitoring between Freestyle Libre Pro and iPro2 in patients with diabetes mellitus. *J Diabetes Investig*. 2019;10(3):851-856.
134. Li A, Riddell MC, Potashner D, Brown RE, Aronson R. Time lag and accuracy of continuous glucose monitoring during high intensity interval training in adults with type 1 diabetes. *Diabetes Technol Ther*. 2019;21(5):286-294.
135. Moser O, Eckstein ML, McCarthy O, et al. Performance of the Freestyle Libre flash glucose monitoring (flash GM) system in individuals with type 1 diabetes: a secondary outcome analysis of a randomized crossover trial. *Diabetes Obes Metab*. 2019;21:2505-2512.
136. Moser O, Pandis M, Aberer F, et al. A head-to-head comparison of personal and professional continuous glucose monitoring systems in people with type 1 diabetes: hypoglycaemia remains the weak spot. *Diabetes Obes Metab*. 2019;21(4):1043-1048.
137. Steineck IIK, Mahmoudi Z, Ranjan A, Schmidt S, Jørgensen JB, Nørgaard K. Comparison of continuous glucose monitoring accuracy between abdominal and upper arm insertion sites. *Diabetes Technol Ther*. 2019;21(5):295-302.
138. Welsh JB, Zhang X, Puhr SA, et al. Performance of a factory-calibrated, teal-time continuous glucose monitoring system in pediatric participants with type 1 diabetes. *J Diabetes Sci Technol*. 2019;13(2):254-258.
139. Zaharieva DP, Riddell MC, Henske J. The accuracy of continuous glucose monitoring and flash glucose monitoring during aerobic exercise in type 1 diabetes. *J Diabetes Sci Technol*. 2019;13(1):140-141.
140. Babaya N, Noso S, Hiromine Y, et al. Flash glucose monitoring in type 1 diabetes: a comparison with self-monitoring blood glucose. *J Diabetes Investig*. 2020;11(5): 1222-1229.
141. Fokkert M, van Dijk P, Edens M, et al. Performance of the Eversense versus the Free Style Libre Flash glucose monitor during exercise and normal daily activities in subjects with type 1 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2020;8(1):e001193.
142. Jafri RZ, Balliro CA, El-Khatib F, et al. A three-way accuracy comparison of the Dexcom G5, Abbott Freestyle Libre Pro, and senseonics eversense continuous glucose monitoring devices in a home-use study of subjects with type 1 diabetes. *Diabetes Technol Ther*. 2020;22(11):846-852.
143. Tsoukas M, Rutkowski J, El-Fathi A, et al. Accuracy of FreeStyle Libre in adults with type 1 diabetes: the effect of sensor age. *Diabetes Technol Ther*. 2020;22(3):203-207.
144. Yan R, Li H, Kong X, et al. The accuracy and precision of the continuously stored data from flash glucose monitoring system in type 2 diabetes patients during standard meal tolerance test. *Int J Endocrinol*. 2020;2020:5947680.
145. Boscari F, Vettoretti M, Amato AML, et al. Comparing the accuracy of transcutaneous sensor and 90-day implantable glucose sensor. *Nutr Metab Cardiovasc Dis*. 2021;31(2): 650-657.
146. Ji L, Guo L, Zhang J, Li Y, Chen Z. Multicenter evaluation study comparing a new factory-calibrated real-time continuous glucose monitoring system to existing flash glucose monitoring system. *J Diabetes Sci Technol*. 2023;17(1):208-213.
147. Link M, Kamecke U, Waldenmaier D, et al. Comparative accuracy analysis of a real-time and an intermittent-scanning continuous glucose monitoring system. *J Diabetes Sci Technol*. 2021;15(2):287-293.
148. Meng R, Gu T, Yang F, Liu J, Sun Q, Zhu D. Performance evaluation of the Glunovo® continuous blood glucose monitoring system in Chinese participants with diabetes: a multicenter, self-controlled trial. *Diabetes Ther*. 2021;12(12):3153-3165.
149. Moser O, Tripolt N, Pferschy P, et al. Performance of the intermittently scanned continuous glucose monitoring (isCGM) system during a high oral glucose challenge in adults with type 1 diabetes: a prospective secondary outcome analysis. *Biosensors (Basel)*. 2021;11(1):22.
150. Nagl K, Berger G, Aberer F, et al. Performance of three different continuous glucose monitoring systems in children with type 1 diabetes during a diabetes summer camp. *Pediatr Diabetes*. 2021;22(2):271-278.
151. 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.
152. Szadkowska A, Michalak A, Łosiewicz A, et al. Impact of factory-calibrated Freestyle Libre System with new glucose algorithm measurement accuracy and clinical performance in children with type 1 diabetes during summer camp. *Pediatr Diabetes*. 2021;22(2):261-270.
153. Alva S, Bailey T, Brazg R, et al. Accuracy of a 14-day factory-calibrated continuous glucose monitoring system with advanced algorithm in pediatric and adult population with diabetes. *J Diabetes Sci Technol*. 2022;16(1):70-77.
154. Boscari F, Vettoretti M, Cavallin F, et al. Implantable and transcutaneous continuous glucose monitoring system: a randomized cross over trial comparing accuracy, efficacy and acceptance. *J Endocrinol Invest*. 2022;45(1):115-124.
155. 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.
156. Garg SK, Liljenquist D, Bode B, et al. Evaluation of accuracy and safety of the next-generation up to 180-day

long-term implantable eversense continuous glucose monitoring system: the PROMISE study. *Diabetes Technol Ther*. 2022;24(2):84-92.
157. Hochfellner DA, Simic A, Taucher MT, et al. Accuracy assessment of the GlucoMen® day CGM system in individuals with type 1 diabetes: a pilot study. *Biosensors*. 2022;12(2):106.
158. Laffel LM, Bailey TS, Christiansen MP, Reid JL, Beck SE. Accuracy of a seventh-generation continuous glucose monitoring system in children and adolescents with type 1 diabetes. *J Diabetes Sci Technol*. 2023;17:962-967.
159. Nakagawa Y, Hirota Y, Yamamoto A, et al. Accuracy of a professional continuous glucose monitoring device in individuals with type 2 diabetes mellitus. *Kobe J Med Sci*. 2022;68(1):E5-E10.
160. Rebec M, Cai K, Dutt-Ballerstadt R, Anderson E. A prospective multicenter clinical performance evaluation of the C-CGM system. *J Diabetes Sci Technol*. 2022;16(2):390-396.
161. Yeoh E, Png D, Khoo J, et al. 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.
162. International Organization for Standardization. In vitro diagnostic test systems: requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus (ISO15197:2013). 2015. https://www.iso.org/standard/54976.html. Accessed July 22, 2023.
163. Food and Drug Administration. *Self—Monitoring Blood Glucose Test Systems for Over—the—Counter Use—Draft Guidance for Industry and Food and Drug Administration Staff*. Silver Spring, MD: Food and Drug Administration; 2014.
164. Pleus S, Eichenlaub M, Gerber T, et al. Improving the bias of comparator methods in analytical performance assessments through recalibration [published online ahead of print October 22, 2022]. *J Diabetes Sci Technol*. doi:10.1177/19322968221133107.
165. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *The Cochrane Database of Systematic Reviews*. 2017;2(2):MR00003.
166. Freckmann G, Baumstark A, Jendrike N, et al. Impact of two different reference neasurement procedures on apparent system accuracy of 18 CE-marked current-generation blood glucose monitoring systems. *J Diabetes Sci Technol*. 2022;16(5):1076-1088.
167. Kölle J, Eichenlaub M, Mende J, et al. Performance assessment of three continuous glucose monitoring systems in adults with type 1 diabetes [published online ahead of print February 22, 2023]. *J Diabetes Sci Technol*. doi:10.1177/19322968231159657.
168. Eichenlaub M, Stephan P, Waldenmaier D, et al. Continuous glucose deviation interval and variability analysis (CG-DIVA): a novel approach for the statistical accuracy assessment of continuous glucose monitoring systems [published online ahead of print November 3, 2022]. *J Diabetes Sci Technol*. doi:10.1177/19322968221134639.
169. Rodbard D. Characterizing accuracy and precision of glucose sensors and meters. *J Diabetes Sci Technol*. 2014;8(5):980-985.
170. Bailey TS. Clinical implications of accuracy measurements of continuous glucose sensors. *Diabetes Technol Ther*. 2017;19(S2):S51-S54.
171. 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.
172. 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.
173. 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.
174. 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.
175. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014;8(4):658-672.
176. 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.
177. Stephan P, Eichenlaub M, Waldenmaier D, et al. A statistical approach for assessing the compliance of integrated continuous glucose monitoring systems with Food and Drug Administration accuracy requirements. *Diabetes Technol Ther*. 2023;25(3):212-216.