

# Evaluating the precision and reliability of real-time continuous glucose monitoring systems in ambulatory settings: a systematic review

Valentina Dávila-Ruales , Laura F. Gilón , Ana M. Gómez , Oscar M. Muñoz, María N. Serrano and Diana C. Henao

## Abstract

**Background:** Continuous glucose monitoring (CGM) with minimally invasive devices plays a key role in the assessment of daily diabetes management by detecting and alerting to potentially dangerous trends in glucose levels, improving quality of life, and treatment adherence. However, there is still uncertainty as to whether CGMs are accurate enough to replace self-monitoring of blood glucose, especially in detecting episodes of hypoglycemia.

**Objectives:** Evaluate clinical, numerical accuracy, sensitivity, and specificity of the CGM devices commercially available when compared to the reference standard of arterial or venous blood glucose.

**Data sources and methods:** We searched the Cochrane Library, PubMed, EMBASE, and LILACS databases. The quality was assessed with the Quality Assessment Diagnostic Accuracy Studies (QUADAS-2) tool. Clinical and numerical accuracy data were extracted. Sensitivity and specificity were calculated using Review Manager software. Heterogeneity was assessed by visual examination of forest plot and summary receiver operating characteristic curves.

**Results:** Twenty-two studies with a total of 2294 patients were included. The average mean absolute relative difference for overall diagnostic accuracy was 9.4%. None of the devices evaluated with ISO 15197:2013 criteria achieved values  $\geq 95\%$  of measurements in the stipulated ranges in hypoglycemia ( $\pm 15$  mg/dL), but two devices did achieve it in hyperglycemia ( $\pm 15\%$ ; Dexcom G6 and G7). Most of the devices evaluated with consensus error grids reached values above 99% in zones A and B only in overall accuracy and hyperglycemia. For hypoglycemia, the average sensitivity was 85.7% and specificity 95.33%, and for hyperglycemia was 97.45% and 96% respectively.

**Conclusion:** Currently available CGM devices have adequate accuracy for euglycemia and hyperglycemia; however, it is still inadequate for hypoglycemia, although it has improved over time.

**Trial registration:** Prospero registration ID CRD42023399767.

**Keywords:** accuracy, continuous glucose monitoring, diabetes mellitus, sensitivity, sensor, specificity

Received: 26 April 2024; revised manuscript accepted: 12 November 2024.

## Introduction

Glucose monitoring is essential to assess glycemic control and optimize treatment. Strategies have evolved from self-blood glucose monitoring

(SBGM) to noninvasive and minimally invasive continuous glucose monitoring (CGM) devices,<sup>1,2</sup> that have opened new horizons in the daily management of diabetes, improving quality of life and

*Ther Adv Endocrinol  
Metab*

2024, Vol. 15: 1–24

DOI: 10.1177/

20420188241304459

© The Author(s), 2024.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:  
**Valentina Dávila-Ruales**  
Department of Internal  
Medicine, Pontificia  
Universidad Javeriana,  
Hospital Universitario San  
Ignacio, Carrera 7 # 40-62,  
Chapinero, Bogotá 110231,  
Colombia  
[v.davila@javeriana.edu.co](mailto:v.davila@javeriana.edu.co)

**Laura F. Gilón**  
**Oscar M. Muñoz**  
Department of Internal  
Medicine, Pontificia  
Universidad Javeriana,  
Hospital Universitario San  
Ignacio, Bogotá, Colombia

**Ana M. Gómez**  
**María N. Serrano**  
**Diana C. Henao**  
Department of Internal  
Medicine, Pontificia  
Universidad Javeriana,  
Hospital Universitario San  
Ignacio, Bogotá, Colombia  
Endocrinology Unit,  
Pontificia Universidad  
Javeriana, Hospital  
Universitario San Ignacio,  
Bogotá, Colombia

adherence to treatment by detecting hypoglycemic and hyperglycemic events not visible by SBGM and providing alerts on potentially dangerous trends in glucose changes, allowing the patient to take preventive measures.<sup>3–5</sup> In addition, a significant reduction in HbA1c has been demonstrated in CGM users in both type 1 diabetes mellitus (TDM1) and type 2 diabetes mellitus (TDM2;  $-0.26\%$  to  $-0.40\%$ ), and the percentage of patients achieving HbA1c  $<7\%$  and  $<8\%$  is higher in CGM users.<sup>6–9</sup> A systematic review and meta-analysis also showed an effect on time-in-range (TIR), with a 5.4% absolute increase in TIR among CGM users.<sup>6</sup>

Karter et al.,<sup>9</sup> showed that hypoglycemia rates decreased from 5.1% to 3.0% among real-time CGM initiators and increased from 1.9% to 2.3% among non-initiators (difference-in-differences estimate:  $-2.7\%$ ; 95% CI:  $-4.4$  to  $-1.1$ ;  $p=0.001$ ), with no statistically significant difference in rates of hyperglycemia or hospitalization for any reason. Similar data were found by Reaven et al.,<sup>8</sup> where CGM initiation was associated with significantly reduced risk of hypoglycemia (hazard ratio (HR): 0.69; 95% CI: 0.48–0.98) in patients with TDM1, however in this study they also find a reduction in all-cause hospitalization in TDM2 and TDM1, HR: 0.75; 95% CI: 0.63–0.90 and HR: 0.89; 95% CI: 0.83–0.97, respectively. As a result, current guidelines recommend CGM as the preferred method of glucose monitoring for all patients with T1DM and T2DM on intensive insulin therapy or at high risk for symptomatic or asymptomatic hypoglycemia.<sup>10,11</sup>

After the first approval of CGM devices by the US Food and Drug Administration (FDA) in 1999, it was cautioned that they could not be widely recommended due to certain limitations, such as the need for multiple calibrations per day, high cost, complications at the sensor insertion site, and inaccuracy of measurements with high false-positive and false-negative rates.<sup>12</sup> Although most of these limitations have been overcome by the new devices available, with a 30% increase in use since 2016,<sup>13</sup> there is still uncertainty as to whether CGMs are accurate enough to replace self-monitoring of blood glucose (SMBG), especially in detecting episodes of hypoglycemia.<sup>14</sup> Teo et al.,<sup>7</sup> found that CGM had no effect on the number of severe hypoglycemic events ( $p=0.13$ ) or diabetic ketoacidosis events ( $p=0.88$ ).

Comparison of the various minimally invasive CGM sensors is complicated by the lack of standardized protocols and methodologies for assessing and reporting CGM accuracy and performance,<sup>15</sup> resulting in a lack of consistency in the metrics reported across studies to assess accuracy.<sup>16–19</sup> Some report sensitivity and specificity to assess accuracy in detecting episodes of hypoglycemia and hyperglycemia, while others report it through numerical accuracy such as MARD (mean absolute relative difference), MAD (mean absolute difference),<sup>20,21</sup> and clinical accuracy measures such as error grids (Clarke, consensus, continuous, surveillance).<sup>22–25</sup> The International Organization for Standardization (ISO) 15197:2013 provides guidance on the criteria that devices must meet. The minimum acceptable criteria are that 95% of glucose monitoring system results are within  $\pm 15$  mg/dL of the values measured by the reference method when glucose concentrations are  $<100$  mg/dL (based on the difference between paired measurements), or within  $\pm 15\%$  when glucose concentrations are  $\geq 100$  mg/dL. For measures of clinical accuracy that describe the probability of making a correct treatment decision based on the assessed test result, 99% of pooled results should fall within zones A and B for the consensus error grid<sup>26</sup> or above 95% for the Clarke grid.<sup>22</sup>

The aim of this systemic review is to evaluate the numerical and clinical accuracy of the different minimally invasive CGM devices currently commercially available in global glycemia, hypoglycemia, and hyperglycemia ranges, as well as to evaluate the sensitivity and specificity for detecting episodes of hypoglycemia and hyperglycemia when compared to a reference standard of venous or arterial blood glucose in patients with T1DM and T2DM.

## Methods

A systematic review of diagnostic test studies was performed. The protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews) ID CRD42023399767.

## Data search and sources

The literature search was performed on December 1 2022 and was updated in July 2024 in the Cochrane Library, PubMed (MedLine), EMBASE (Elsevier), and LILACs databases,

restricting the search to studies published in English and Spanish and date from January 1, 2018, to July 1, 2024. Search terms can be found in Supplemental Material (Supplement 1).

### *Study selection*

We included prospective studies in adults and/or children with T1DM or T2DM that evaluated the numerical accuracy, clinical accuracy, sensitivity, and specificity of minimally invasive CGM sensors for detecting hypoglycemia or hyperglycemia events compared with the reference standard (venous or arterial blood glucose). Gestational diabetes, cystic fibrosis-related diabetes, studies in exercise, critically ill or hospitalized patients, and those in which the reference test was exclusively capillary glucose were excluded. In vitro studies and those performed in species other than humans also were excluded.

The sensors included were those commercially available in 2024: FreeStyle Libre 2 (Abbott Diabetes Care, Alameda, CA, USA), FreeStyle Libre 3 (Abbott Diabetes Care, Alameda, CA, USA), Eversense (Senseonics, Inc., Germantown, MD, USA), Eversense XL (Senseonics, Inc., Germantown, MD, USA), Guardian Sensor 3 (Medtronic Diabetes, Northridge, CA, USA), Dexcom G4 Platinum (Dexcom Inc., San Diego, CA, USA), Dexcom G6 (Dexcom Inc., San Diego, CA, USA), Dexcom G7 (Dexcom Inc., San Diego, CA, USA), AiDEX (MicroTech Medical (Hangzhou) Co. Ltd., Zhejiang, China), GlucoMen (WaveForm Diabetes, Wilsonville, OR, USA), Glunovo (Infinovo, Suzhou, China), A6 TouchCare (Medtrum Technologies, Inc., Shanghai, China), CareSens Air (I-sens, Inc., Incheon, South Korea), and SiJoy System (Sibionics Shenzhen Technology Co. Ltd., China).

Trials had to report at least one (1) clinical accuracy outcome according to ISO 15197:2013 standards (Clarke, continuous, consensus, or surveillance error grids), (2) numerical accuracy outcomes (MARD, MAD, percentage of index test results within  $\pm 15$  or  $\pm 20$  mg/dL of reference method values for glucose concentrations  $< 100$ ,  $< 80$ , or  $< 70$  mg/dL, percentage of results within  $\pm 15\%$  or  $\pm 20\%$  for glucose concentrations  $\geq 100$ ,  $\geq 80$ , or  $\geq 70$  mg/dL), or (3) evaluation of operational characteristics (sensitivity and specificity) for detection of hypoglycemia  $\leq 70$  mg/dL and hyperglycemia  $> 180$  mg/dL. Studies that

defined different thresholds for the diagnosis of hypoglycemia or hyperglycemia were also included. When more than one reference standard was reported, only the information from the venous and/or arterial blood reference test was considered.

### *Definition of the minimum acceptable criteria that must be met by the devices*

The ISO 15197:2013 provides guidance on the criteria that devices must meet. The minimum acceptable criteria are that 95% of glucose monitoring system results are within  $\pm 15$  mg/dL of the values measured by the reference method when glucose concentrations are  $< 100$  mg/dL (based on the difference between paired measurements), or within  $\pm 15\%$  when glucose concentrations are  $\geq 100$  mg/dL. For measures of clinical accuracy that describe the probability of making a correct treatment decision based on the assessed test result, 99% of pooled results should fall within zones A and B for the consensus error grid<sup>26</sup> or above 95% for the Clarke grid.<sup>22</sup>

### *Data extraction and quality assessment*

Two reviewers (VD, LG) selected studies in parallel and independently, first on the basis of the title and abstract and then on the basis of the full text. Disagreements were resolved by consensus or with the participation of a third investigator (OM, NS). Two investigators independently assessed the quality of the included studies (VD, LG) using the QUADAS-2 tool (Quality Assessment Diagnostic Accuracy Studies)<sup>27</sup> and classified each study as “low,” “high,” or “unclear” risk of bias and as “low,” “high,” or “unclear” concerns about applicability.

### *Data synthesis and analysis*

Data extraction was done in a paired way. The data extracted for each of the studies were: design, author, year, country, number of subjects and participating centers, baseline characteristics (age, sex, BMI, HbA1c, type of diabetes, and number of child and adult participants), number of paired samples analyzed, glucose threshold, and characteristics of the reference test (laboratory technique used and time interval between sensor measurements and reference test). MARD and MAD data were obtained for overall glycemia, and individually in ranges of hyperglycemia,

euglycemia, and hypoglycemia. When studies evaluated different thresholds, the main analysis was performed using the thresholds for hypoglycemia ( $\leq 70$  mg/dL) and hyperglycemia ( $> 180$  mg/dL) defined by the American Diabetes Association.<sup>28</sup>

The information was summarized in comparative tables for each of the numerical and clinical accuracy outcomes in the areas of overall glycemia, hypoglycemia, and hyperglycemia.

The sensitivity and specificity of each sensor, and the respective 95% confidence intervals were calculated using the Review Manager software (RevMan 5.4<sup>®</sup>) proposed by The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen.<sup>29</sup> The hierarchical summary receiver operating characteristic (SROC) curves were drawn using RevMan 5.4.

Heterogeneity was assessed by visual inspection of the SROC curves and the forest plot, as suggested by the Cochrane Collaboration.<sup>30</sup>  $I^2$  statistic was not used as it does not account for heterogeneity explained by phenomena such as positivity threshold effects.<sup>30</sup>

#### Publication bias

The funnel plot test has low power to detect publication bias in studies of diagnostic tests when there is considerable heterogeneity, so it was not performed in the present study.<sup>31</sup>

## Results

The PRISMA flow diagram shows the selection process. Of 7984 studies initially identified, 22 were finally included in the analysis<sup>32–53</sup> (Figure 1). The reasons for the exclusion of the remaining studies are reported in Supplemental Material (Supplement 2). In total, the selected studies included 2294 patients and 320,216 paired measurements.

The characteristics of the included studies are shown in Table 1. Eight studies included pediatric populations<sup>33,37,38,40,41,46,50,51</sup> and two did not evaluate adults.<sup>40,46</sup> Four studies evaluated only patients with T1DM.<sup>36,37,40,49</sup> The reference test in most studies was venous or arterial blood glucose measured by Yellow Spring Instruments (YSI; 2300 STAT Plus Glucose and Lactate

Analyzer).<sup>32–46,49–53</sup> Nine studies used devices that do not require calibration.<sup>33,38,40–42,45,46,51,53</sup> The sensor with the longest duration of use was the Eversense XL at 180 days.<sup>37,39</sup>

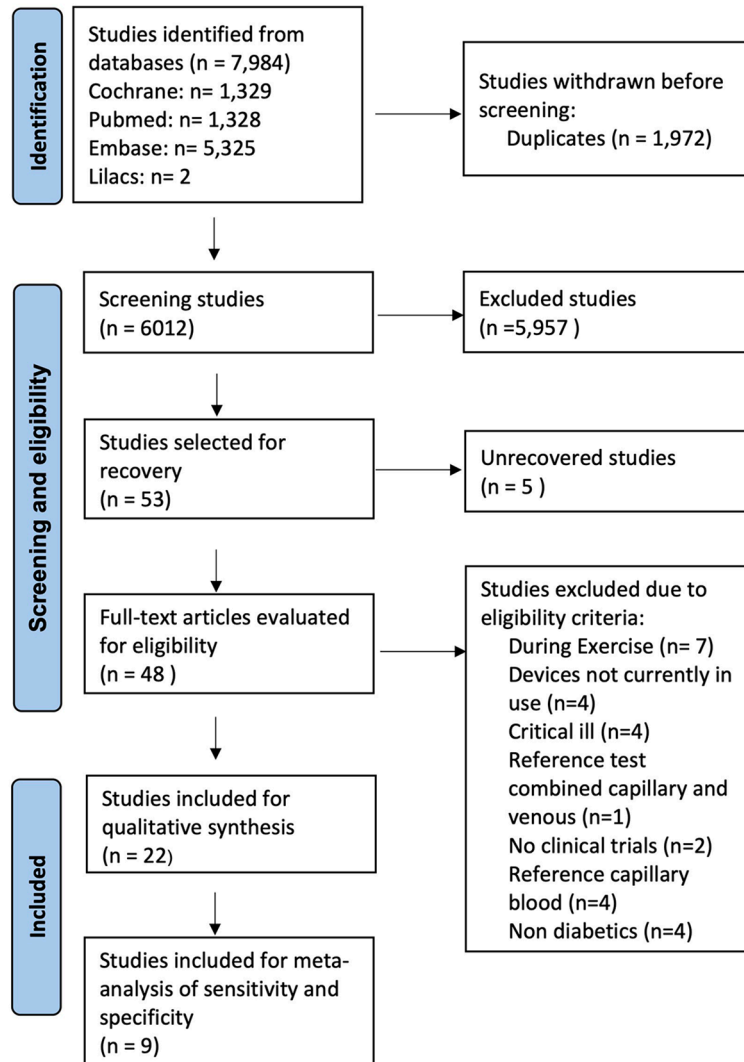
The quality assessment of the studies is shown in Figure 2. In general, the highest risk of bias was found for flow and timing, because not all patients were included in the analysis without stating the reason, nor was there an adequate interval between the index test and the reference standard.<sup>32–34,36,38–40,45,47,51–53</sup> Concerns about applicability in general were low because all studies included patients with T1DM, T2DM, and the reference standard with venous blood, and most measured with Yellow Spring Instruments 2300 (YSI, Inc., Yellow Springs, OH, USA), which is widely accepted by most manufacturers as a method for reference measurements and device calibration,<sup>54</sup> two studies used Entwicklung, Konstruktion und Fertigung (EKF) as the venous blood measurement.<sup>47,48</sup>

#### Overall diagnostic accuracy

Most studies assessed accuracy in the 40–500 mg/dL range (Table 2). MARD was the most used metric to assess numerical device accuracy across the entire glycemic range. The average MARD was 9.4%, with the best value being 7.7% (Dexcom G6),<sup>38,40</sup> followed by 7.8% (Free Style Libre 3).<sup>51</sup> Six studies reported MARD  $> 10\%$ .<sup>36,43,48,50,52,53</sup>

Only nine studies<sup>32,37,38,40–42,46,49,52</sup> were based on ISO 15197:2013 standards,<sup>26</sup> however, none reached values higher than 95% of the measurements in the stipulated ranges ( $\pm 15$  mg/dL or  $\pm 15\%$ ). According to this parameter, the device with the highest accuracy was Dexcom G6 with 91.1%,<sup>38,40</sup> and the lowest was Eversense (Senseonics) with 65.6%<sup>49</sup> (Table 2).

Eleven studies evaluated consensus or Parkes error grids,<sup>32,33,37,43–45,47,48,50–52</sup> 10 of which found values in zone A + B greater than 99%.<sup>32,33,37,43,45,47,48,50–52</sup> Nine reported Clarke's error grids.<sup>32,34,36,38,45,47–50</sup> Only one reported value below 95% in zones A + B.<sup>49</sup> One study reported Clarke's measurements of 0.04% in zone E,<sup>48</sup> five studies reported in zone D, four of which were below 1%,<sup>32,34,45,47,50</sup> except one which reported percentages of 2.4%<sup>36</sup> (Table 3). Five studies additionally reported error grids that



**Figure 1.** PRISMA flow diagram showing selection process.

are not part of the ISO 15197:2013 criteria (continuous and surveillance)<sup>32,37,38,40,53</sup> (Table 3).

### *Diagnostic accuracy in hypoglycemia*

The included studies used thresholds between 40 and 80 mg/dL (Table 4), with some reporting different thresholds simultaneously<sup>34,35,39,41–44,46,48,51–53</sup>. Only nine studies evaluated accuracy in terms of MARD for hypoglycemia.<sup>32,34,35,37,39,40,47,48,53</sup> The sensor with the best MARD was FreeStyle Libre 3 with 3.6% for the <54 mg/dL range,<sup>53</sup> followed by Eversense (Senseonics) with 7.2% for the 40–60 mg/dL range.<sup>35</sup> The device with the highest MARD was G7 System<sup>53</sup> (53.4% at <54 mg/dL and 27% for the 55–69 mg/dL threshold),

followed by Glunovo (35% at <54 mg/dL and 19.7% for the 55–69 mg/dL threshold).<sup>48</sup>

No device reached the ISO 15197:2013 recommendations for the percentage of measurements in the recommended range (>95% of measurements in the  $\pm 15$  mg/dL range), with values below 40% reported for A6 TouchCare<sup>32</sup> (Table 4). Using broader accuracy criteria than those stipulated by ISO ( $\pm 20$  mg/dL for reference values  $\leq 70$ –100 mg/dL, or  $\pm 20\%$  for reference values  $> 70$ –100 mg/dL) three devices exceeded 95% for different hypoglycemia thresholds, FreeStyle Libre 3,<sup>51,53</sup> FreeStyle Libre 2,<sup>33</sup> Eversense (updated algorithm),<sup>35</sup> and G7 Dexcom.<sup>42,46</sup> Nevertheless, the study comparing



**Table 1.** Characteristics of the included studies.

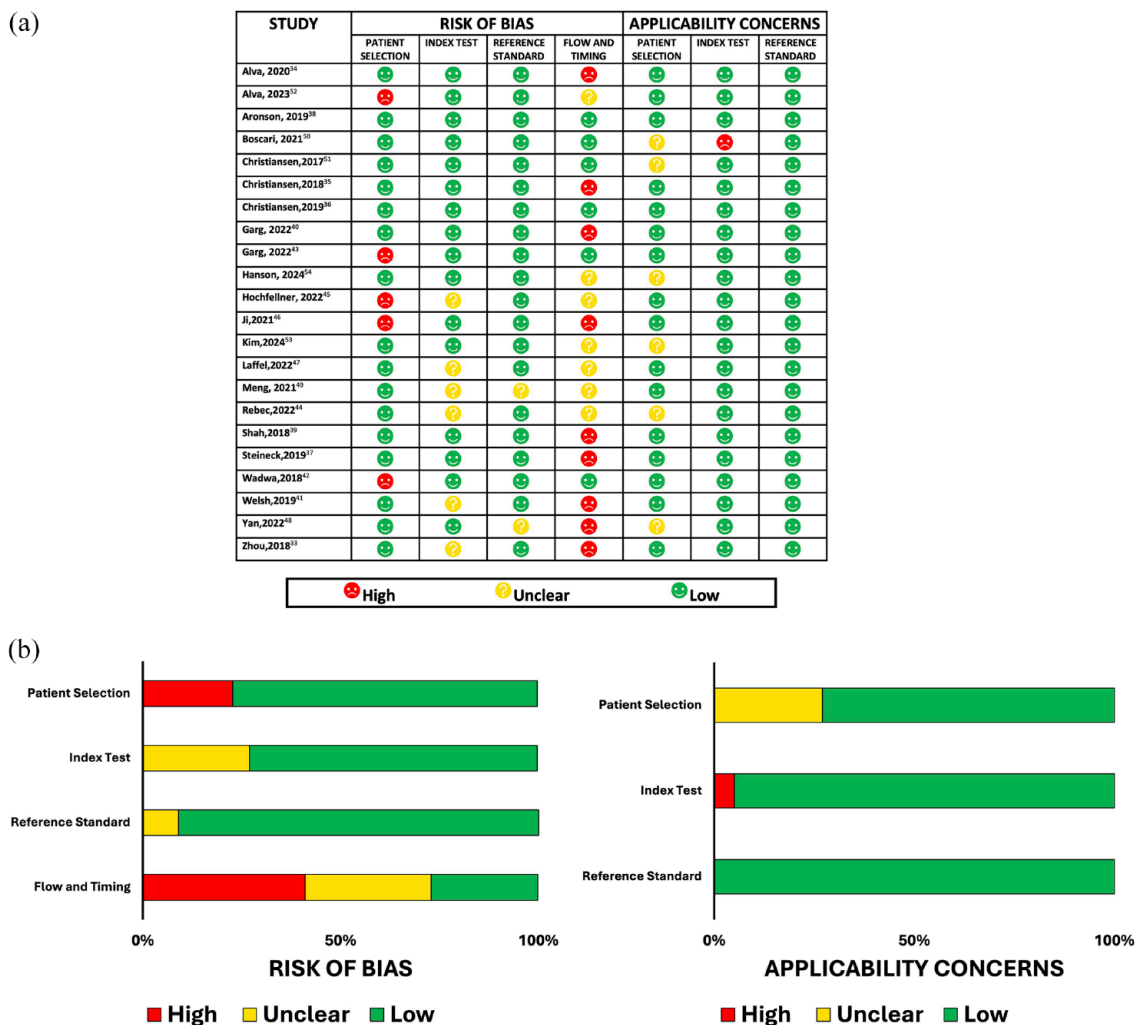
Author, year	Device	Days of use	Insertion/calibration per day	Country	Male, %	Mean BMI, kg/m <sup>2</sup>	Mean HbA1c, %	Number of participants/paired measurements	Reference test	Insulin challenge	Type of diabetes (%)	Mean age (years)/inclusion of children
Alva, 2020 <sup>33</sup>	FreeStyle Libre 2 (Abbott)	14	Arm/0 (Factory-calibrated)	USA	46.6 A 55.4 C	28.3 A 21 C	7.8 A 8.3 C	144/18,926 A 129/6584 C	YSI 2300	Yes	T1DM: 91.1 A T2DM: 9.9 A T1DM: 98.6 C T2DM: 1.4 C	47.3/yes 12.2 C
Alva, 2023 <sup>51</sup>	FreeStyle Libre 3 (Abbott)	14	Arm/0 (Factory-calibrated)	USA	44	29.6 A 23.1 C	7.4 A 8.5 C	56/4769 A 39/2076 C	YSI 2300	No	T1DM: 83 T2DM: 17	49.7/yes 13.3 C
Aronson, 2019 <sup>37</sup>	Eversense XL (Senseonics)	180	Arm/2	Canada	64	22.3	8	36/7163	YSI 2300	No	T1DM: 100	16.9/Yes
Boscarl, 2021 <sup>49</sup>	Eversense (Senseonics)	90	NR/2	Italy	81.8	NR	7.5	11/388	YSI 2300	Yes	T1DM: 100	47.4/No
Christiansen, 2017 <sup>50</sup>	Guardian Sensor 3 sensor (Medtronic)	7	Arm-Abdomen/2-4	USA	52.3	28.2	7.9	88/11,619	YSI 2300	Yes	T1DM: 70.5 T2DM: 29.6	42/Yes
Christiansen, 2018 <sup>54</sup>	Eversense (Senseonics)	90	Arm/2	USA	54	29.1	7.6	90/16,653	YSI 2300	Yes	T1DM: 61 T2DM: 29	45.1/No
Christiansen, 2019 <sup>35</sup>	Eversense-Updated algorithm	90	Arms/2	USA	51.4	28.2	NR	35/15,170	YSI 2300	Yes	T1DM: 71.4 T2DM: 28.6	51.6/No
Garg, 2022 <sup>39</sup>	Eversense (Senseonics)	180	Arm/2	USA	47	31.4	7.6	208/49,613	YSI 2300	Yes	T1DM: 69.6 T2DM: 30.4	48.6/No
Garg, 2022 <sup>42</sup>	G7 System (Dexcom)	10	Arm-abdomen/0 (Factory-calibrated)	USA	46.5	28.9	NR	318/77,774	YSI 2300	Yes	T1DM: 80.8 T2DM: 19.1	44.3/No
Hanson, 2024 <sup>53</sup>	FreeStyle Libre 3 (Abbott)	14	Arm/0 (Factory-calibrated)	USA	39.3	30.3	7.5	56/4020	YSI 2300	No	TDM1: 58.9 TDM2: 41.1	49.9/No
Hanson, 2024 <sup>53</sup>	G7 system (Dexcom)	10	Arm/0 (Factory-calibrated)	USA	39.3	30.3	7.5	56/3640	YSI 2300	No	TDM1: 58.9 TDM2: 41.1	49.9/No
Hochleitner, 2022 <sup>44</sup>	Glucomen (WaveForm Cascade)	14	Abdomen/1	Austria	62.5	28	7	8/450	YSI 2300	Yes	NR	41.6/No
Ji, 2021 <sup>45</sup>	AlDEX (Microtech Medical)	14	Arm-Abdomen/0 (Factory-calibrated)	China	49.6	25.5	7.4	120/14,586	YSI 2300	No	T1DM: 11.3 T2DM: 88.7	60.2/No
Kim, 2024 <sup>52</sup>	CareSens Air (I-sens)	15	Arm/1	Korea	45.2	24.6	7.8	84/10,029	YSI 2300	No	T1DM: 75 T2DM: 25	40.1/No

(Continued)

Table 1. (Continued)

Author, year	Device	Days of use	Insertion/ calibration per day	Country	Male, %	Mean BMI, kg/m <sup>2</sup>	Mean HbA1c, %	Number of participants/ paired measurements	Reference test	Insulin challenge	Type of diabetes (%)	Mean age (years)/ inclusion of children
Laffel, 2022 <sup>46</sup>	G7 System (Dexcom)	10	Arm-Abdomen/0 (Factory-calibrated)	USA	52.4	NR	NR	164/115,437	YSI 2300	Yes	T1DM: 100	12.2 (no adult)
Meng, 2021 <sup>48</sup>	Gluconovo (Infinovo)	14	Abdomen/2	China	52.56	NR	NR	78/12,688	EKF	NR	T1DM: 32 T2DM: 68	NR/No
Rebec, 2022 <sup>43</sup>	GlucoMen (WaveForm Cascade)	14	Abdomen/1	Slovenia, Croatia, Serbia	47	25.8	7.4	60/17,823	YSI 2300	Yes	T1DM: 84 T2DM: 16	46/No
Shah, 2018 <sup>38</sup>	G6 System (Dexcom)	10	Arm-Buttocks/1 (Factory-calibrated)	USA	61	NR	8.2 A 8.1 C	76/3532	YSI 2300	No	T1DM: 96 T2DM: 4	25.5/Yes
Steineck, 2019 <sup>36</sup>	G4 Platinum (Dexcom)	4	Abdomen/2	Denmark	57.1	26	7	14/2660	YSI 2300	Yes	T1DM: 100	48/No
Wadwa, 2018 <sup>41</sup>	G6 System (Dexcom)	10	Abdomen-Buttock/1 (Factory-calibrated)	USA	47	NR	8	290/21,560	YSI 2300	Yes	T1DM: 99.2 T2DM: 0.8	28/Yes
Welsh, 2019 <sup>40</sup>	G6 System (Dexcom)	10	NR/0 (Factory-calibrated)	USA	39	NR	8.1	49/1378	YSI 2300	No	T1DM: 100	13.5 (no adult)
Yan, 2022 <sup>47</sup>	SiJoy System (Sibionics)	14	Arm/NR	China	42	23.1	NR	78/NR	EKF	NR	T1DM: 56.5 T2DM: 34.8	41.5/No
Zhou, 2018 <sup>32</sup>	A6 TouchCare (Medtrum)	7	Arm/2	China	55.5	24.7	8.2	63/1678	YSI 2300	Yes	T1DM: 16 T2DM: 84	59/No

A, adults; C, children; EKF, Entwicklung, Konstruktion und Fertigung, blood glucose/lactate analyzers; NR, not reported; T1DM, type 1 diabetes; T2DM, type 2 diabetes; YSI 2300, STAT Plus Glucose and Lactate Analyzer (YSI, Inc., Yellow Springs, OH, USA), using the glucose-oxidase method.



**Figure 2.** (a) Quality assessment with the QUADAS-2 tool: Risk of bias and applicability concerns. Both were assessed on three key domains: Patient selection, index test, reference standard, and a fourth domain (flow and timing) was assessed only for risk of bias. (b) Risk of bias graph.

Patient selection domain: (1) Risk of bias in five studies was high<sup>41,42,44,45,51</sup> considering that the operational variables for inclusion and exclusion criteria were not adequately assessed or the exclusion and inclusion criteria were not documented. In two studies the exclusion criteria were inappropriate by not specifying which were those conducts that could be negatively impacted by glucose manipulation or medical conditions that could compromise patient or staff safety, which were selected at the discretion of the investigator,<sup>41,42</sup> and three did not name which were the exclusion criteria in the study.<sup>44,45,51</sup> (2) The applicability in six studies<sup>43,47,49,50,52,53</sup> was uncertain considering that many exclusion criteria were used. Domain index test: (1) Uncertain risk of bias in six studies,<sup>32,40,43,44,46,48</sup> since they did not report the time interval between the paired measurements of the CGM under study and that of the standard reference, so it is not possible to know if the readings of the sensor under study were interpreted without knowing the results of the reference test. (2) Applicability: Only one study<sup>49</sup> generated doubts in its applicability since the accuracy was evaluated 2 months after the sensor was inserted and the accuracy was evaluated only in 7 days of the 90 days of the sensor duration. Domain reference standard: (1) Uncertain risk of bias in two studies<sup>47,48</sup> since the measurement of central glycemia was different from YSI. (2) Applicability in all studies was rated as adequate. Flow and timing domain: (1) Most studies had high risk of bias<sup>32-34,36,38-40,45,47</sup> or uncertain risk of bias<sup>43,44,46,48,51-53</sup> because there was not an adequate interval between the index test and the reference test as well as not including all patients in the analysis. CGM, continuous glucose monitoring; QUADAS-2, Quality Assessment Diagnostic Accuracy Studies; YSI, Yellow Spring Instruments.



**Table 2.** Overall accuracy of CGM devices.

Study	Population/site insertion	Threshold (mg/dL)	MARD% (95% CI)	±20/20%	±15/15%
Alva, 2020/FreeStyle Libre 2 (Abbott) <sup>33</sup>	Adults	40–500	9.2 (8.7–9.9)	93.2 <sup>‡</sup>	86.3 <sup>§</sup>
	Children	40–500	9.7 (8.9–10.7)	92.1 <sup>‡</sup>	85.5 <sup>§</sup>
Alva, 2023/FreeStyle Libre 3 (Abbott) <sup>51</sup>	Overall	40–500	7.8	93.4 <sup>  </sup>	NR
	Children	40–500	8.6	89.7 <sup>  </sup>	NR
	Adults	40–500	7.5	94.9 <sup>  </sup>	NR
Aronson, 2019/Eversense XL (Senseonics) <sup>37</sup>					
90 days	Overall	40–400	9.1 (8.8–9.4)	NR	NR
180 days	Overall	40–400	9.4 (8.6–10.5)	NR	83.4 <sup>§</sup>
	Children	40–400	9.7 (8.6–10.8)	NR	NR
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	Overall	NR	NR	NR	65.6 <sup>§</sup>
Christiansen, 2017/Guardian Sensor 3 (Medtronic) <sup>50</sup>					
MiniMed 640G-Abdomen					
Minimum calibration	Overall	NR	10.6 (10.4–10.7)	88.2 <sup>‡</sup>	NR
Additional calibration*	Overall	NR	9.6 (9.4–9.8)	90.7 <sup>‡</sup>	NR
Christiansen, 2018/Eversense (Senseonics) <sup>34</sup>	Overall	40–400	8.8 (8.1–9.3)	93.3	85.7
Christiansen, 2019/Eversense (Senseonics) <sup>35</sup>					
Updated algorithm	Overall	40–400	9.6 (8.9–10.4)	93 <sup>‡</sup>	85 <sup>§</sup>
Updated algorithm PRECISE II	Overall	40–400	8.5 (8–9.1)	94 <sup>‡</sup>	87 <sup>§</sup>
Garg, 2022/Eversense (Senseonics) <sup>39</sup>	Overall	40–400	9.1 (9.0–9.2)	92.9 <sup>‡</sup>	85.6 <sup>§</sup>
	SBA*	Overall	40–400	8.5 (8.4–8.6)	93.9 <sup>‡</sup>
Garg, 2022/G7 System (Dexcom) <sup>42</sup>	Arm	40–400	8.2	95.3*	89.6 <sup>§</sup>
	Abdomen	40–400	9.1	93.2*	85.5 <sup>§</sup>
Hanson, 2024/FreeStyle Libre 3 (Abbott) <sup>53</sup>	Overall	NR	9.8	91.4 <sup>  </sup>	85 <sup>  </sup>
Hanson, 2024/G7 System (Dexcom) <sup>53</sup>	Overall	NR	13.3	78.6 <sup>  </sup>	64.7 <sup>  </sup>
Hochfellner, 2022/GlucoMen (WaveForm) <sup>44</sup>	Overall	100–400	9.7(8.9–10.6)	NR	NR
Ji, 2021/AiDEX (Microtech Medical) <sup>45</sup>	Overall	NR	9.0	95	86
Kim, 2024/CareSens Air (I-sens) <sup>52</sup>	Overall	40–500	10.4	89*	78.5 <sup>§</sup>
Laffel, 2022/G7 System (Dexcom) <sup>46</sup>	Arm	80–300	8.1	95.3*	88.8 <sup>§</sup>
	Abdomen	80–300	9	92.9*	86 <sup>§</sup>
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	Overall	NR	10.3 (9.5–11)	89.71 <sup>‡</sup>	79.3 <sup>§</sup>

*(Continued)*

**Table 2.** (Continued)

Study	Population/site insertion	Threshold (mg/dL)	MARD% (95% CI)	±20/20%	±15/15%
Rebec, 2022/GlucoMen (WaveForm) <sup>43</sup>					
Regular algorithm	Overall	100–400	11.5	80.5	68.4
Hybrid algorithm	Overall	100–400	9.9	85	74
Shah, 2018/G6 System (Dexcom) <sup>38</sup>					
	Overall	40–400	9 (7.9–10.1)	93.9*	83.3 <sup>§</sup>
	Adults	40–400	9.8 (8.1–11.5)	92.5*	78.3 <sup>§</sup>
	Children	40–400	7.7 (6.6–8.8)	96.2*	91.1 <sup>§</sup>
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>					
	Abdomen	NR	12.3 (11.5–12.7)	NR	NR
	Arm	NR	12 (11.5–12.5)	NR	NR
Wadwa, 2018/G6 System (Dexcom) <sup>41</sup>					
	Overall	40–400	10 (9.6–10.4)	92.3*	82.4 <sup>§</sup>
	Adults	40–400	9.9 (9.4–10.4)	92.4*	82.6 <sup>§</sup>
	Children	40–400	10.1 (9.2–11)	91.9*	81.6 <sup>§</sup>
Welsh, 2019/G6 System (Dexcom) <sup>40</sup>					
	Overall	40–400	7.7	96.2*	91.1 <sup>§</sup>
Yan, 2022/SiJoy System (Sibionics) <sup>47</sup>					
	Overall	NR	8.8 (8.6–8.9)	91.8 <sup>‡</sup>	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>					
	Overall	40–400	9.1 (8.9–9.2)	90.5*	81.5 <sup>§</sup>

Minimum calibration two times per day, Additional calibration three to four times per day. Update algorithm application of the updated glucose calculation algorithm to the PRECISE II sensor.

\*20% or ±20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ±20% for glucose concentrations ≥100 mg/dL.

§15% or ±15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ±15% for glucose concentrations ≥100 mg/dL.

‡20% or ±20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ±20% for glucose concentrations ≥80 mg/dL.

§15% or ±15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ±15% for glucose concentrations ≥80 mg/dL.

||20% or ±20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±20% for glucose concentrations ≥70 mg/dL.

||15% or ±15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±15% for glucose concentrations ≥70 mg/dL.

CGM, continuous glucose monitoring; MARD, mean absolute relative difference; NR, not reported; SBA, sacrificial boronic acid sensor, specific modification to glucose-binding indicator chemistry, improving longevity by reducing oxidation.

Freestyle Libre 3 and Dexcom G7 reported values for the Dexcom G7 as low as 0% for the <54 mg/dL threshold and 67.7% for the 55–69 mg/dL threshold, but it should be noted that they used a glucose concentration at 20% or ±20 mg/dL of 70 mg/dL.<sup>53</sup>

None of the five studies that evaluated error grids met the risk zone parameter in A and B above 99% for consensus, neither above 95% for Clarke<sup>32,36,47–49</sup> (Table 5). Only one study reported

the percentage of measurements in other risk zones with the G4 Platinum Device (Dexcom),<sup>36</sup> finding that 0% were in zone E, but between 23% (measurements in the arm) and 27% (measurements in the abdomen) were in zone D.

#### Diagnostic accuracy in hyperglycemia

The accuracy thresholds evaluated were 180–400 mg/dL (Table 6). Sixteen studies evaluated MARD<sup>32,34–41,44–47,51–53</sup> with values <10%, except

**Table 3.** Clinical accuracy error grids for overall glycemia.

Study	Clarke A/B (A + B) %	Consensus A/B (A + B) %	Continuous A/B (A + B) %	Surveillance none—slight, lower %
Alva, 2020/FreeStyle Libre 2 (Abbott) <sup>33</sup>	NR	93.2/6.7 (99.9) A	NR	NR
	NR	92.6/7.4 (100) C	NR	NR
Alva, 2023/FreeStyle Libre 3 (Abbott) <sup>51</sup>	NR	92.1/7.8 (99.9)	NR	NR
Aronson, 2019/Eversense XL (Senseonics) <sup>37</sup>				
90 days	NR	93.4/6.2 (99.6)	NR	NR
180 days	NR	NR	NR	98.4
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	77.6/16.7 (94.3)	NR	NR	NR
Christiansen, 2017/Guardian Sensor 3 (Medtronic) <sup>50</sup>				
MiniMed 640G-Abdomen				
Minimum calibration	87.7/11.4 (99.1)	87.7/12.2 (99.9)	NR	NR
Additional calibration*	90.1/9.1 (99.2)	90.3/9.6 (99.9)	NR	NR
Christiansen, 2018/Eversense (Senseonics) <sup>34</sup>	92.8/6.5 (99.3)	NR	NR	NR
Hanson, 2024/FreeStyle Libre 3 (Abbott) <sup>53</sup>	NR	NR	NR	99
Hanson, 2024/G7 System (Dexcom) <sup>53</sup>	NR	NR	NR	98.2
Hochfellner, 2022/GlucoMen (WaveForm) <sup>44</sup>	NR	84.9/12.9 (97.8)	NR	NR
Ji, 2021/AiDEX (Microtech Medical) <sup>45</sup>	95.7/4.2 (99.9)	95/5 (100)	NR	NR
Kim, 2024/CareSens Air (I-sens) <sup>52</sup>	NR	89.9/10 (99.9)	NR	NR
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	89/10 (99)	89.1/10.7 (99.8)	NR	NR
Rebec, 2022/GlucoMen (WaveForm) <sup>43</sup>				
Regular algorithm	NR	89.3/10 (99.3)	NR	NR
Hybrid algorithm	NR	91.4/8 (99.4)	NR	NR
Shah, 2018/G6 System (Dexcom) <sup>38</sup>	91.9/8 (99.9) A	NR	NR	99.3 A
	95.7/4.1 (99.8) C	NR	NR	NR
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>	85.6/11.6 (97.2) Abd	NR	NR	NR
	86/11.6 (97.6) Ar	NR	NR	NR
Welsh, 2019/G6 System (Dexcom) <sup>40</sup>	NR	NR	NR	99.6
Yan, 2022/SiJoy System (Sibionics) <sup>47</sup>	89.8/9.4 (99.2)	97.6/2.3 (99.9)	NR	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>	89.7/9.4 (99.1)	94.3/5.5 (99.8)	71.8/19 (90.8)	97.9
Minimum calibration two times per day, Additional calibration three to four times per day. A, adults; Abd, abdomen insertion; Ar, arm insertion; C, children; NR not reported.				

**Table 4.** Hypoglycemia accuracy of CGM devices.

Study	Population/ site insertion	Threshold (mg/dL)	MARD% (95% CI)	MAD (mg/dL; 95% CI)	±20/20%	±15/15%
Alva, 2020/FreeStyle Libre 2 (Abbott) <sup>33</sup>	Adults	70	NR	NR	98.4 <sup>‡</sup>	94.3 <sup>§</sup>
	Children	70	NR		98.8 <sup>‡</sup>	96.1 <sup>§</sup>
Alva, 2023/FreeStyle Libre 3(Abbott) <sup>51</sup>	Overall	<54	NR	16.1	80 <sup>  </sup>	66.7 <sup>  </sup>
		54–69	NR	1.1	95.2 <sup>  </sup>	88.6 <sup>  </sup>
Aronson, 2019 /Eversense XL (Senseonics) <sup>37</sup>						
90 days	Overall	70	10.5 (9.2–11.8)	NR	NR	NR
	Children	70	10.6 (9–12.1)	NR	NR	NR
	Adults	70	10.3 (7.9–12.7)	NR	NR	NR
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	Overall	70	NR	NR	NR	68.3 <sup>§</sup>
Christiansen, 2017/Guardian Sensor 3 (Medtronic) <sup>50</sup>						
MiniMed 640G-Abdomen						
Minimum calibration	Overall	70	NR	NR	92.5 <sup>‡</sup>	NR
Additional calibration*	Overall	70	NR	NR	92.8 <sup>‡</sup>	NR
Christiansen, 2018/Eversense (Senseonics) <sup>34</sup>	Overall	40–54	10.7 (7.4–13.3)	NR	85.6	83.2
		55–69	9 (8.1–12)	NR	92.9	86.1
Christiansen, 2019/Eversense (Senseonics) <sup>35</sup>						
Updated algorithm	Overall	40–60	7.2	NR	96 <sup>‡</sup>	92 <sup>§</sup>
		61–80	7.6	NR	94 <sup>‡</sup>	87 <sup>§</sup>
Updated algorithm PRECISE II	Overall	40–60	8.3	NR	92 <sup>‡</sup>	85 <sup>§</sup>
		61–80	8.7	NR	91 <sup>‡</sup>	83 <sup>§</sup>
Garg, 2022/Eversense (Senseonics) <sup>39</sup>	Overall	40–60	9.4 (9.1–9.7)	NR	89.4 <sup>‡</sup>	83.2 <sup>§</sup>
		61–80	8.8 (8.6–9)	NR	92.2 <sup>‡</sup>	84.1 <sup>§</sup>
Garg, 2022/G7 System (Dexcom) <sup>42</sup>	Arm	40–60	NR	8.5	91*	85.1 <sup>§</sup>
		61–80	NR	6.3	96.5*	92.6 <sup>§</sup>
	Abdomen	40–60	NR	10.3	85*	77.1 <sup>§</sup>
		61–80	NR	7.3	94.1*	89.4 <sup>§</sup>
Hanson, 2024/FreeStyle Libre 3 (Abbott) <sup>53</sup>	Overall	<54	3.6	NR	100 <sup>  </sup>	NR
		55–69	13.7	NR	88.2 <sup>  </sup>	NR
Hanson, 2024/G7 System (Dexcom) <sup>53</sup>	Overall	<54	53.4	NR	0 <sup>  </sup>	NR
		55–69	27	NR	67.7 <sup>  </sup>	NR

(Continued)

**Table 4.** (Continued)

Study	Population/ site insertion	Threshold (mg/dL)	MARD% (95% CI)	MAD (mg/dL; 95% CI)	±20/20%	±15/15%
Hochfellner, 2022/GlucoMen (WaveForm) <sup>44</sup>	Overall	40–70	NR	19.5	NR	NR
Ji, 2021/AiDEX (Microtech Medical) <sup>45</sup>	Overall	70	NR	9.7	93	86.7
Kim, 2024/CareSens Air (I-sens) <sup>52</sup>	Overall	<54	NR	11.8	88.2	76.4
		54–69	NR	12.4	77.6*	67.1 <sup>§</sup>
Laffel, 2022/G7 System (Dexcom) <sup>46</sup>	Arm	40–60	NR	11.3	85.3*	74.4 <sup>§</sup>
		61–80	NR	6.4	95.5*	93 <sup>§</sup>
	Abdomen	40–60	NR	15.6	73.1*	56 <sup>§</sup>
		61–80	NR	9	90.6*	85.9 <sup>§</sup>
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	Overall	<54	35 (8.7–61.4)	NR	42.3 <sup>‡</sup>	34.6 <sup>§</sup>
		54–69	19.7 (13.7–25.8)	NR	66.7 <sup>‡</sup>	52 <sup>§</sup>
Rebec, 2022/GlucoMen (WaveForm) <sup>43</sup>						
Regular Algorithm	Overall	40–50	NR	NR	61.7	51.3
		50–80	NR	NR	69.1	57.2
Hybrid Algorithm	Overall	40–50	NR	NR	57.1	48.4
		50–80	NR	NR	71	58.9
Shah, 2018/G6 system (Dexcom) <sup>38</sup>	Overall	70	NR	9.5	90.8*	80 <sup>§</sup>
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>	Abdomen	70	NR	15.1 (13–17)	NR	NR
	Arm	70	NR	16 (14–18)	NR	NR
Wadwa, 2018/G6 System (Dexcom) <sup>41</sup>	Overall	<54	NR	10.9	85*	78.7 <sup>§</sup>
		54–69	NR	7.8	94.4*	89.5 <sup>§</sup>
Welsh, 2019/G6 System (Dexcom) <sup>40</sup>	Overall	70	13.3	9.1	92.6*	81.5 <sup>§</sup>
Yan, 2022/SiJoy System (Sibionics) <sup>47</sup>	Overall	80	15 (13.6–16.5)	NR	82.9 <sup>‡</sup>	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>	Overall	70	16.6	12.7	72*	36 <sup>§</sup>

Minimum calibration two times per day, Additional calibration three to four times per day. Update algorithm application of the updated glucose calculation algorithm to the PRECISE II sensor.

\*20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ±20% for glucose concentrations ≥100 mg/dL.

§15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ±15% for glucose concentrations ≥100 mg/dL.

‡20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ±20% for glucose concentrations ≥80 mg/dL.

§15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ±15% for glucose concentrations ≥80 mg/dL.

||20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±20% for glucose concentrations ≥70 mg/dL.

¶15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±15% for glucose concentrations ≥70 mg/dL.

MARD, mean absolute relative difference; MAD, mean absolute difference in mg/dL; NR, not reported.

**Table 5.** Summary of studies evaluating clinical accuracy by error grids in hypoglycemia.

Study	Clarke A + B %	Consensus A + B %	Surveillance none—slight, lower %
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	65.8	NR	NR
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	40.3	96.9	NR
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>	73 Abd	NR	NR
	76.6 Ar	NR	NR
Yan, 2022/SiJoy System (Sibionics) <sup>47</sup>	88	98.2	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>	NR	96	96
A, adults; Abd, abdomen insertion; Ar, arm insertion; C, children; NR, not reported.			

for Glunovo, CareSens Air at the threshold of >250 mg/dL with a MARD of 10.1% and 10.4%, respectively.<sup>48,52</sup>

Only two devices met the ISO 15197:2013 recommended criteria for hyperglycemia (>95% of measurements in the  $\pm 15\%$  or 15 mg/dL range), Dexcom G6 at the >250 mg/dL threshold<sup>40</sup> and Dexcom G7 in arm insertion at the 301–400 mg/dL threshold.<sup>42,46</sup> Using wider ranges ( $\pm 20$  mg/dL or 20%), two devices were above 95% (Dexcom G7 and Dexcom G6)<sup>40,42,46</sup> and two met it only in the >250 mg/dL range<sup>38,41</sup> (Table 6).

Five studies evaluated clinical accuracy in terms of error grids<sup>32,36,47–49</sup> (Table 7). All the evaluated devices met the expected parameters.

#### *Sensitivity and specificity for hyperglycemia and hypoglycemia events*

Nine studies provided sufficient data to estimate sensitivity and specificity for hypoglycemia and hyperglycemia events.<sup>33–36,39,41,42,45,47</sup>

Figure 3 shows the forest plot of sensitivity and specificity for the detection of hypoglycemic events. For the 70 mg/dL threshold, the mean sensitivity was 85.7% and specificity was 95.33%. For the 60 mg/dL threshold, the values were 84%

and 97%, respectively. Sensitivity was lower at lower thresholds. There is significant heterogeneity in sensitivity, but not in specificity.

Figure 4 shows the forest plot of sensitivity and specificity for the detection of hyperglycemic events. For the 180 mg/dL threshold, the average sensitivity was 97.45% and specificity was 96%. For the 200 mg/dL threshold, 93.4% and 89.6%, respectively. For the 240 mg/dL threshold, the mean sensitivity was 94.4% and specificity was 98.4%. Graphically, there is no significant heterogeneity.

#### *Adverse events*

Sixteen studies reported adverse events.<sup>32,34,35,37–39,42,43,45–48,50–53</sup> Most reported an adverse event rate of less than 15%, and two studies reported no device-related adverse events.<sup>43,47</sup> In terms of frequency, the most reported adverse events were dermatologic (erythema, edema, induration, skin irritation associated with the adhesive patch, skin atrophy, and hypopigmentation), followed by hematologic (bruising, bleeding) and pain during device insertion or sensor removal. One study<sup>37</sup> reported presyncope, nausea, and vomiting in six patients associated with device insertion or removal. Four studies reported bleeding,<sup>39,45,51,53</sup> which was considered mild.



**Table 6.** Accuracy in hyperglycemia of CGM devices.

Study	Population/site insertion	Threshold (mg/dL)	MARD% (95% CI)	±20/20%	±15/15%
Alva, 2022/FreeStyle Libre 2 (Abbott) <sup>33</sup>	Adults	180	NR	95 <sup>‡</sup>	89.6 <sup>§</sup>
	Children	180	NR	95.7 <sup>‡</sup>	88.7 <sup>§</sup>
Alva, 2023/FreeStyle Libre 3 (Abbott) <sup>51</sup>	Overall	181–250	6.3	95.8 <sup>  </sup>	92.4 <sup>  </sup>
		>250	4.9	100 <sup>  </sup>	98.2 <sup>  </sup>
Aronson, 2019/Eversense XL (Senseonics) <sup>37</sup>					
90 days	Overall	180	6.7 (6.3–7.1)	NR	NR
	Children	180	6.8 (6.3–7.3)	NR	NR
	Adults	180	6.6 (5.6–7.7)	NR	NR
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	Overall	180	NR	NR	90 <sup>§</sup>
Christiansen, 2017/Guardian Sensor 3 (Medtronic) <sup>50</sup>					
MiniMed 640G-Abdomen					
Minimum calibration	Overall	180	NR	89.2 <sup>‡</sup>	NR
Additional calibration*	Overall	180	NR	93.5 <sup>‡</sup>	NR
Christiansen, 2018/Eversense (Senseonics) <sup>34</sup>	Overall	180	7.8 (7.2–8.8)	95	86.9
Christiansen, 2019/Eversense (Senseonics) <sup>35</sup>					
Updated algorithm	Overall	181–300	8.6	93 <sup>‡</sup>	85 <sup>§</sup>
		301–350	6.9	98 <sup>‡</sup>	93 <sup>§</sup>
		351–400	6.4	96 <sup>‡</sup>	92 <sup>§</sup>
Updated algorithm PRECISE II	Overall	181–300	7.8	96 <sup>‡</sup>	88 <sup>§</sup>
		301–350	7	98 <sup>‡</sup>	91 <sup>§</sup>
		351–400	5.2	99 <sup>‡</sup>	97 <sup>§</sup>
Garg, 2022/Eversense (Senseonics) <sup>39</sup>	Overall	181–300	7.7 (7.6–7.8)	94.7 <sup>‡</sup>	87.9 <sup>§</sup>
		301–350	7.1 (7–7.2)	96.5 <sup>‡</sup>	90.6 <sup>§</sup>
		351–400	8 (7.7–8.3)	93.9 <sup>‡</sup>	87.8 <sup>§</sup>
Garg, 2022/G7 System (Dexcom) <sup>42</sup>	Arm	181–300	NR	96 <sup>*</sup>	90.3 <sup>§</sup>
		300–400	NR	99.1 <sup>*</sup>	96.8 <sup>§</sup>
	Abdomen	181–300	NR	93.4 <sup>*</sup>	85.1 <sup>§</sup>
		300–400	NR	98.6 <sup>*</sup>	93.5 <sup>§</sup>
Hanson, 2024/FreeStyle Libre 3 (Abbott) <sup>53</sup>	Overall	181–250	8.3	93.3 <sup>  </sup>	NR
		>250	7.5	96.3 <sup>  </sup>	NR
Hanson, 2024/G7 System (Dexcom) <sup>53</sup>	Overall	181–250	10.8	86.9 <sup>  </sup>	NR
		>250	10.6	93.3 <sup>  </sup>	NR
Hochfellner, 2022/GlucoMen (WaveForm) <sup>44</sup>	Overall	201–400	6.1	NR	NR

*(Continued)*

Table 6. (Continued)

Study	Population/site insertion	Threshold (mg/dL)	MARD% (95% CI)	±20/20%	±15/15%
Ji, 2021/AiDEX (Microtech Medical) <sup>45</sup>	Overall	180	8.7	NR	NR
Kim, 2024/CareSens Air (I-sens) <sup>52</sup>	Overall	181–250	9.5	92.6*	81.5 <sup>§</sup>
		>250	10.4	87*	74.8 <sup>§</sup>
Laffel, 2022/G7 System (Dexcom) <sup>46</sup>	Arm	181–300	7.6	97*	88.5 <sup>§</sup>
		301–400	5.4	99.4*	96.9 <sup>§</sup>
	Abdomen	181–300	7.1	97.4*	90.4 <sup>§</sup>
		301–400	5.7	99.6*	95.4 <sup>§</sup>
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	Overall	180–250	9.9	92.4 <sup>‡</sup>	81.4 <sup>§</sup>
		>250	10.1	91.6 <sup>‡</sup>	83.7 <sup>§</sup>
Rebec, 2022/GlucoMen (WaveForm) <sup>43</sup>					
Regular algorithm	Overall	181–300	NR	85.3	72.7
		300–400	NR	75.2	60.9
Hybrid algorithm	Overall	181–300	NR	91.4	80.7
		300–400	NR	86.4	74.2
Shah, 2018/G6 System (Dexcom) <sup>38</sup>	Overall	181–250	8.9	92.9*	82.6 <sup>§</sup>
		>250	6.3	96.2*	92.3 <sup>§</sup>
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>	Abdomen	180	9.6 (8.9–10.3)	NR	NR
	Arm	180	6.3 (5.8–6.9)	NR	NR
Wadwa, 2018/G6 System (Dexcom) <sup>41</sup>	Overall	181–250	9.2	92.4*	80.8 <sup>§</sup>
		>250	7.2	97.4*	90.7 <sup>§</sup>
Welsh, 2019/ G6 System (Dexcom) <sup>40</sup>	Overall	181–250	7.7	95.1*	89.8 <sup>§</sup>
		>250	4.5	100*	97.7 <sup>§</sup>
Yan, 2022/SiJoy SYSTEM (Sibionics) <sup>47</sup>	Overall	200	8.6 (7.9–9.3)	91.5 <sup>‡</sup>	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>	Overall	180	8.1	93*	85.9 <sup>§</sup>

Minimum calibration two times per day, Additional calibration three to four times per day. Update algorithm application of the updated glucose calculation algorithm to the PRECISE II sensor.  
 \*20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ±20% for glucose concentrations ≥100 mg/dL.  
<sup>§</sup>15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <100 mg/dL or within ± 15% for glucose concentrations ≥100 mg/dL.  
<sup>‡</sup>20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ±20% for glucose concentrations ≥80 mg/dL.  
<sup>§</sup>15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <80 mg/dL or within ± 15% for glucose concentrations ≥80 mg/dL.  
<sup>‡</sup>20% or ± 20 mg/dL, which was the percentage of sensor values that fell within either ±20 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±20% for glucose concentrations ≥70 mg/dL.  
<sup>§</sup>15% or ± 15 mg/dL, which was the percentage of sensor values that fell within either ±15 mg/dL of the reference values for glucose concentrations <70 mg/dL or within ±15% for glucose concentrations ≥70 mg/dL.  
 CGM, continuous glucose monitoring; MARD, mean absolute relative difference; NR, not reported.

**Table 7.** Summary of studies evaluating clinical accuracy by error grids in hyperglycemia.

Study	Clarke A + B %	Consensus A + B %	Surveillance none—slight, lower %
Boscari, 2021/Eversense (Senseonics) <sup>49</sup>	100	NR	NR
Meng, 2021/Gluconovo (Infinovo) <sup>48</sup>	99.6	99.8	NR
Steineck, 2019/G4 Platinum (Dexcom) <sup>36</sup>	100 Abd	NR	NR
	100 Ar	NR	NR
Yan, 2022/SiJoy System (Sibionics) <sup>47</sup>	99.8	100	NR
Zhou, 2018/A6 TouchCare (Medtrum) <sup>32</sup>	NR	99.8	100

A, adults; Abd, abdomen insertion; Ar, arm insertion; C, children; NR not reported.

Only one serious event was reported, associated with the inability to remove the sensor on the first attempt, requiring surgical intervention.<sup>34</sup>

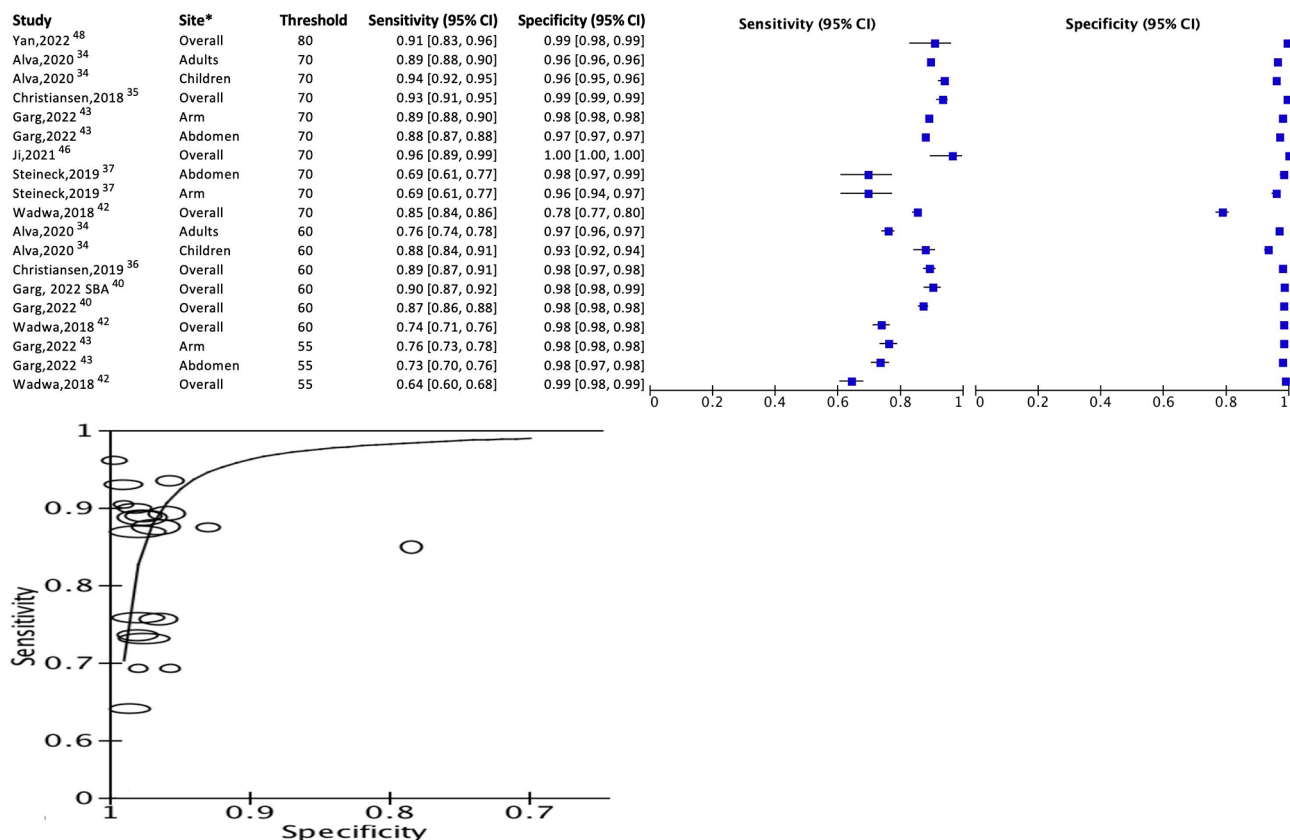
### Discussion

Our systematic review suggests that the accuracy of the various commercially available CGM sensors is adequate in the range of global glycemia and hyperglycemia, both in terms of numerical accuracy and clinical accuracy as measured by error grids. However, accuracy is still limited in hypoglycemia ranges, which could limit clinical decisions based on sensor measurements. The findings are similar in terms of sensitivity and specificity for detecting hypoglycemia events.

MARD is the most widely used numerical accuracy parameter due to its ease of interpretation, with an accepted cut-off point of <10%–12%.<sup>20,21,55</sup> We found that most of the sensors reported a MARD <10%, for both overall and hyperglycemia ranges. However, it was evident that MARD varies according to the characteristics of the assessment, tending to be lower when measurements are made on the arm compared to the abdomen, and in adults relative to children. In addition, the differences between the parameters used to assess precision (e.g., glucose ranges, varying rates of changing glucose, and day of sensor wear), it has been documented that the first day of performance is usually worse than the remaining days.<sup>56</sup> Other potential sources of heterogeneity not evaluated in our study include calibration differences between sensors, calibration errors,<sup>16,17,57,58</sup> manufacturing batches, and time

of assessment. Discrepancies in reported accuracy statistics are to be expected because of the lack of standardized protocols and methodologies for assessing and reporting CGM accuracy and performance.<sup>15</sup> Therefore, MARD may be influenced by multiple factors beyond sensor performance.<sup>16</sup> This is why ISO 15197:2013 proposes different criteria, such as the proportion of measurements within a specified range relative to a reference method >95%.<sup>26</sup> We found that only a relatively small proportion of studies reported ISO criteria and none of the sensors met this parameter when evaluating the global or hypoglycemia range, but two devices (G6 and G7 Dexcom) did meet it in the range of hyperglycemia >250 mg/dL.<sup>40,42,46</sup> These results highlight the importance of complete and standardized reporting for the new devices to avoid reporting bias.

In terms of clinical accuracy criteria, most studies in the global and hyperglycemic ranges reported that the percentage of measurements in risk zones A + B for Clarke and consensus was greater than 95% and 99%, respectively, suggesting that no errors would be made in clinical decisions based on CGM measurements. It is noteworthy, however, that there was very limited reporting of the percentages in the other risk zones, where measurement errors would lead to undesirable clinical outcomes. The limited information available suggests that approximately 1% of sensor readings could lead to misinterpretation, although the risk of adverse clinical outcomes is unclear. In addition, the sensitivity and specificity for assessing hyperglycemia were good for most sensors



**Figure 3.** Sensitivity and specificity for detection of hypoglycemia. (a) Forrest plot; (b) Hierarchical SROC curve.

\*Site of insertion of the sensor and population.

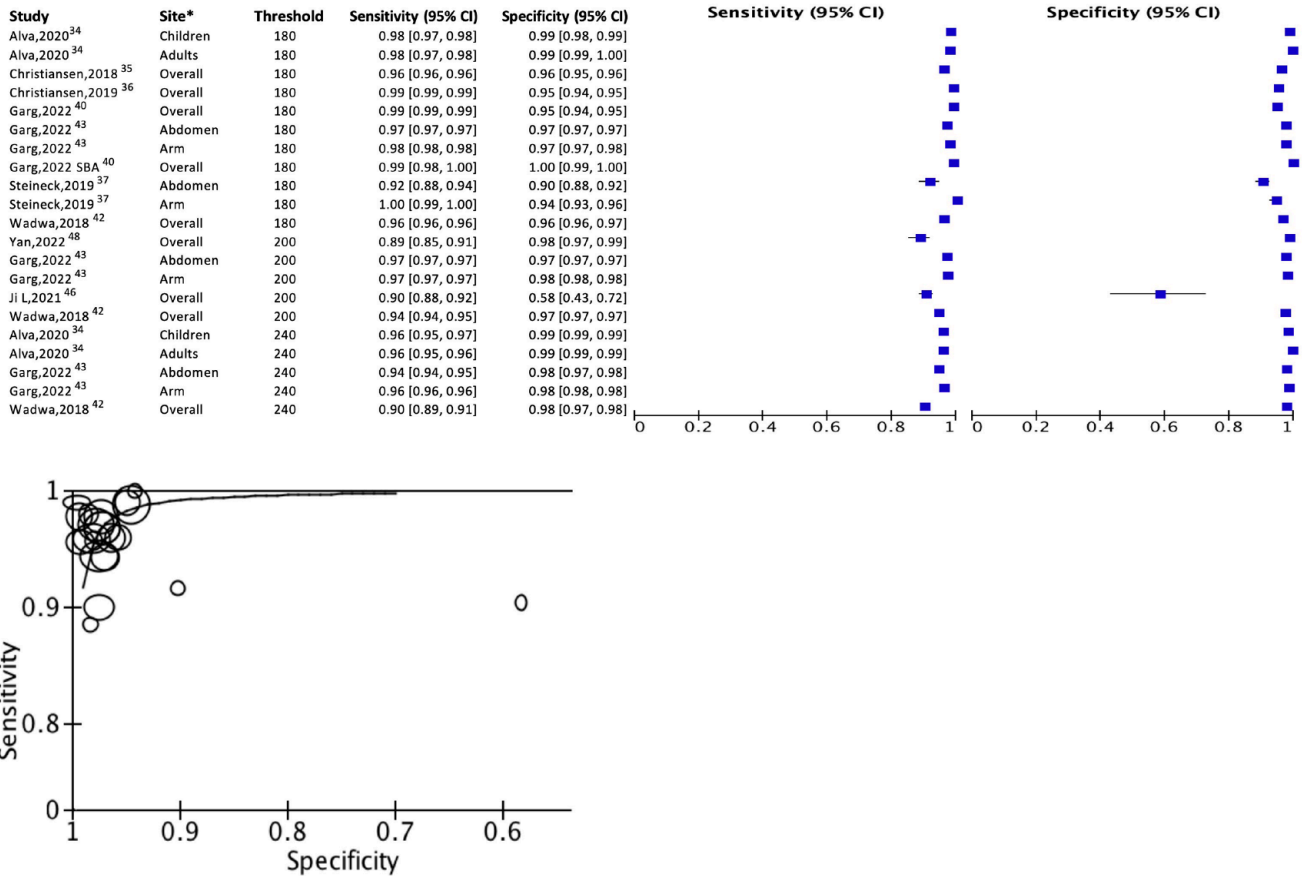
SROC, summary receiver operating characteristic.

(97.45% and 96%, respectively). Thus, most sensors have sufficient clinical accuracy and are safe for making treatment decisions in this range.

The data presented in this study demonstrate that current sensors continue to have suboptimal operating characteristics for the diagnosis of hypoglycemic events, with an average sensitivity for glycemia <70 mg/dL of 85.7% and <60 mg/dL of 84%, but with good specificity (95.3% and 97%, respectively). A recent systematic review and meta-analysis that evaluated the diagnostic accuracy of different sensors for detecting hypoglycemia in T1DM and T2DM found an average sensitivity and specificity for detecting hypoglycemia much lower than our study (69.3% and 93.3%, respectively), with a high frequency of false-positive and false-negative alarms.<sup>14</sup> This difference is due to the fact that this meta-analysis

included sensors with older technology, some of which are no longer commercially available.

We additionally evaluated other precision parameters in hypoglycemia ranges. MARD was variable with values as high as 53.4%–35% at the glucose threshold of <54 mg/dL<sup>48,53</sup> but also with some <10%.<sup>34,35,39,53</sup> Nevertheless, it is known that MARD estimates are subject to relatively large errors in the hypoglycemic range, in part due to a markedly nonlinear relationship with glucose level, and also when there are only a small number of observations in the hypoglycemic range, especially at the lower end of the hypoglycemic range, one might expect to obtain a MARD value that is closer to the observed values in the target range,<sup>17</sup> as seen in the study where the MARD for Dexcom G7 was 53. Four percent for the <54 mg/dL range and 27% for the



**Figure 4.** Sensitivity and specificity for detection of hyperglycemia. (a) Forrester plot; (b) Hierarchical SROC curve. \*Site of insertion of the sensor and population. SROC, summary receiver operating characteristic.

54–69 mg/dL range,<sup>53</sup> but this may be explained by the fact that the numbers in this hypoglycemic range were only 1 and 27, respectively.

None of the devices met the ISO 15197:2013 recommended parameters for the proportion of measurements in a specific range, but three devices exceeded the 95% threshold for measurements with a higher acceptance range ( $\pm 20$  mg/dL for reference values  $\leq 80$ – $100$  mg/dL or  $\pm 20\%$  for reference values  $> 80$ – $100$  mg/dL).<sup>33,35,42,46</sup> None of the devices achieved a percentage of measurements in risk zones A + B of the error grids greater than 95% in Clarke and more than 99% in consensus, but three studies<sup>32,47,48</sup> reported percentages of measurements in these zones greater than 95% for the consensus grid. In addition, one study reported that there were no measurements in zone E,<sup>36</sup> which would represent the highest risk zone for inaccurate

measurements, suggesting that the percentage of clinical conduct where errors can be made is low.<sup>59</sup> This demonstrates the progress in technology in hypoglycemia ranges, but there are still limitations in terms of safety.

With more sensitive sensors, different algorithms have been developed to improve the accuracy and reliability of the devices, allowing the development of closed-loop systems or “artificial pancreas” that automatically pump insulin according to the predicted value of the glucose level and avoid dangerous glycemic states.<sup>2,60,61</sup> Such devices also have the ability to detect trends, fluctuations, and rapid changes throughout the day, providing hypoglycemia alerts that can be used by the patient to take rapid preventive action. The high number of readings, taken every 1–5 min,<sup>3–5</sup> could reduce the number of false positives and false negatives.

With the data available to date, it is recommended that all values be confirmed at hypoglycemic thresholds to avoid false alarms. In addition, it is recommended to obtain capillary recordings in the presence of hypoglycemic symptoms to avoid false negatives. However, the benefits of CGM in terms of reduced HbA1c, fewer severe hypoglycemic events, increased TIR, reduced hospitalizations for severe hypoglycemia, and diabetic ketoacidosis outweigh these limitations.<sup>6,8,9,62–67</sup> Several controlled clinical trials have demonstrated the large benefit of hypoglycemia reduction, Haak et al.<sup>62</sup> showing a 43% reduction in hypoglycemia for glycemia <70 mg/dL, 53% for <55 mg/dL, and 64% for events <45 mg/dL, with a significant improvement in patient satisfaction. A recent meta-analysis found that CGM patients spend less time in hypoglycemia than SBGM patients.<sup>68</sup> In addition, SBGM is an invasive and uncomfortable procedure for the patient as it requires a digital puncture, which means that adherence can be as low as 24%–44% for T2DM and T1DM, respectively,<sup>69,70</sup> therefore, CGM is being positioned as the standard for glucose monitoring in people with diabetes to achieve better adherence and therefore better glycemic control.<sup>11,71,72</sup>

The present systematic review shows a low percentage of adverse events, most of which were mild dermatologic reactions. Only one study reported a serious but nonfatal event,<sup>34</sup> which may improve patient compliance, in addition to other advantages of current devices such as longer sensor life in the body, fewer calibrations, and even factory calibration.

Our study has several strengths. We evaluated current and new FDA-approved devices available on the market, which gives an up-to-date view of the state of the technology. Additionally, we evaluated multiple accuracy metrics in addition to sensitivity and specificity, covering not only the hypoglycemia range but also global glycemia and hyperglycemia thresholds. Finally, we compared the devices only to the reference standard of venous or arterial blood.

However, there are limitations that need to be recognized. Accuracy metrics are not fully standardized, leading to heterogeneity in the reporting of primary studies. In addition, only a limited number of studies report results using the criteria

proposed by ISO, which introduces a risk of publication bias. Similarly, only a proportion of studies report clinical accuracy metrics, and these data are particularly limited at the hypoglycemic threshold. We, therefore, insist on the need to standardize measurements and reporting of this type of study. Finally, we excluded from the review studies that evaluated patients hospitalized in general wards or intensive care units, so our results are not generalizable to these populations, where conditions such as hypoperfusion, vasoactive and inotropic support, certain medications, and uremia may alter the accuracy and performance of the devices.<sup>73–75</sup>

### Conclusion

Current sensors available for CGM have adequate accuracy in the overall and hyperglycemia range. For hypoglycemia, the accuracy of the latest sensors on the market has improved but is still low. Until a sufficiently high accuracy is achieved according to standardized requirements, it is still necessary to confirm hypoglycemia levels with capillary blood.

### Declarations

#### *Ethics approval and consent to participate*

The study and the protocol were approved by the Research and Ethics Committee of the Pontificia Universidad Javeriana and the Hospital Universitario San Ignacio in Bogotá. Consent to participate: Not applicable.

#### *Consent for publication*

Not applicable.

#### *Author contributions*

**Valentina Dávila-Ruales:** Investigation; Resources; Writing – original draft; Writing – review & editing.

**Laura F. Gilón:** Investigation; Writing – original draft; Writing – review & editing.

**Ana M. Gómez:** Conceptualization; Investigation.

**Oscar M. Muñoz:** Investigation; Methodology.

**María N. Serrano:** Investigation; Resources.

**Diana C. Henao:** Formal analysis; Supervision.



### Acknowledgements

None.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Competing interests

The authors declare that there is no conflict of interest.

### Availability of data and materials

Not applicable.

### ORCID iDs

Valentina Dávila-Ruales  <https://orcid.org/0000-0001-6563-2939>

Laura Gilón  <https://orcid.org/0000-0002-8089-1987>

Ana M. Gómez  <https://orcid.org/0000-0002-8907-3470>

### Supplemental material

Supplemental material for this article is available online.

### References

- Clarke SF and Foster JR. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. *Br J Biomed Sci* 2012; 69: 83–93.
- Villena Gonzales W, Mobashsher AT and Abbosh A. The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors (Basel)* 2019; 19: 800.
- Cengiz E and Tamborlane WV. A tale of two compartments: interstitial versus blood glucose monitoring. *Diabetes Technol Ther* 2009; 11(Suppl. 1): S11–S116.
- 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: 105–113.
- Thennadil SN, Rennert JL, Wenzel BJ, et al. Comparison of glucose concentration in interstitial fluid, and capillary and venous blood during rapid changes in blood glucose levels. *Diabetes Technol Ther* 2001; 3: 357–365.
- Elbalschy 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: e14854.
- Teo E, Hassan N, Tam W, et al. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. *Diabetologia* 2022; 65: 604–619.
- Reaven PD, Newell M, Rivas S, et al. Initiation of continuous glucose monitoring is linked to improved glycemic control and fewer clinical events in type 1 and type 2 diabetes in the Veterans Health Administration. *Diabetes Care* 2023; 46: 854–863.
- Karter AJ, Parker MM, Moffet HH, et al. Association of real-time continuous glucose monitoring with glycemic control and acute metabolic events among patients with insulin-treated diabetes. *JAMA* 2021; 325: 2273.
- Blonde L, Umpierrez GE, Reddy SS, et al. American Association of Clinical Endocrinology clinical practice guideline: developing a diabetes mellitus comprehensive care plan—2022 Update. *Endocr Pract* 2022; 28: 923–1049.
- ElSayed NA, Aleppo G, Bannuru RR, et al. American Diabetes Association Professional Practice Committee; 6. Glycemic goals and hypoglycemia: standards of care in diabetes—2024. *Diabetes Care* 2024; 47(Suppl. 1): S111–S125.
- Wadwa RP, Fiallo-Scharer R, Vanderwel B, et al. Continuous glucose monitoring in youth with type 1 diabetes. *Diabetes Technol Ther* 2009; 11(Suppl. 1): S83–S91.
- Foster NC, Miller KM, Dimeglio L, et al. Marked increases in CGM use has not prevented increases in HbA1c levels in participants in the T1D exchange (T1DX) clinic network. *Diabetes* 2018; 67(Suppl. 1): 1689-P.
- Lindner N, Kuwabara A and Holt T. Non-invasive and minimally invasive glucose monitoring devices: a systematic review and meta-analysis on diagnostic accuracy of hypoglycaemia detection. *Syst Rev* 2021; 10: 145.
- Freckmann G, Eichenlaub M, Waldenmaier D, et al. Clinical performance evaluation of continuous glucose monitoring systems: a scoping review and recommendations for reporting. *J Diabetes Sci Technol* 2023; 17: 1506–1526.

16. 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: 1030–1040.
17. Rodbard D. Characterizing accuracy and precision of glucose sensors and meters. *J Diabetes Sci Technol* 2014; 8: 980–985.
18. D’Archangelo MJ. New guideline supports the development and evaluation of continuous interstitial glucose monitoring devices. *J Diabetes Sci Technol* 2008; 2: 332–334.
19. Damiano ER, McKeon K, El-Khatib FH, et al. A comparative effectiveness analysis of three continuous glucose monitors: the Navigator, G4 Platinum, and Enlite. *J Diabetes Sci Technol* 2014; 8: 699–708.
20. Vashist SK. Continuous glucose monitoring systems: a review. *Diagnostics (Basel)* 2013; 3: 385–412.
21. Reiterer F, Polterauer P, Schoemaker M, et al. Significance and reliability of MARD for the accuracy of CGM systems. *J Diabetes Sci Technol* 2017; 11: 59–67.
22. Clarke WL. The original Clarke Error Grid Analysis (EGA). *Diabetes Technol Ther* 2005; 7: 776–779.
23. Pfützner A, Klonoff DC, Pardo S, et al. Technical aspects of the Parkes error grid. *J Diabetes Sci Technol* 2013; 7: 1275–1281.
24. Clarke W, Anderson S and Kovatchev B. Evaluating clinical accuracy of continuous glucose monitoring systems: continuous glucose–error grid analysis (CG-EGA). *Curr Diabetes Rev* 2008; 4: 193–199.
25. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol* 2014; 8: 658–672.
26. ISO 15197:2013. In vitro diagnostic test systems—requirements for blood glucose monitoring systems for self-testing in managing diabetes mellitus, <https://www.iso.org/standard/54976.html> (2013, accessed 12 December 2022).
27. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–536.
28. Seaquist ER, Anderson J, Childs B, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care* 2013; 36: 1384–1395.
29. Review Manager (RevMan) [Computer program]. Version 5.4.1, <http://revman.cochrane.org> (2020, accessed 9 January 2023).
30. Deeks JJ, Bossuyt PM, Leeflang MM, et al. (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2.0 (updated July 2023). Cochrane, <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current> (2023, accessed July 2024)
31. van Enst WA, Ochodo E, Scholten RJPM, et al. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol* 2014; 14: 70.
32. 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: 286–293.
33. 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: 70–77.
34. 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.
35. Christiansen MP, Klaff LJ, Bailey TS, et al. 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.
36. Steineck IIK, Mahmoudi Z, Ranjan A, et al. Comparison of continuous glucose monitoring accuracy between abdominal and upper arm insertion sites. *Diabetes Technol Ther* 2019; 21: 295–302.
37. Aronson R, Abitbol A and 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: 1689–1694.
38. Shah VN, Laffel LM, Wadwa RP, et al. Performance of a factory-calibrated real-time continuous glucose monitoring system utilizing an automated sensor applicator. *Diabetes Technol Ther* 2018; 20: 428–433.

39. 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: 84–92.
40. Welsh JB, Zhang X, Pühr SA, et al. Performance of a factory-calibrated, real-time continuous glucose monitoring system in pediatric participants with type 1 diabetes. *J Diabetes Sci Technol* 2019; 13: 254–258.
41. Wadwa RP, Laffel LM, Shah VN, et al. 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: 395–402.
42. 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: 373–380.
43. Rebec M, Cai K, Dutt-Ballerstadt R, et al. A prospective multicenter clinical performance evaluation of the C-CGM system. *J Diabetes Sci Technol* 2022; 16: 390–396.
44. 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 (Basel)* 2022; 12: 106.
45. Ji L, Guo L, Zhang J, et al. 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: 208–213.
46. Laffel LM, Bailey TS, Christiansen MP, et al. Accuracy of a seventh-generation continuous glucose monitoring system in children and adolescents with type 1 diabetes. *J Diabetes Sci Technol* 2023; 17: 262–267.
47. Yan L, Li Q, Guan Q, et al. Evaluation of the performance and usability of a novel continuous glucose monitoring system. *Int J Diabetes Dev Countries* 2023; 43: 551–558.
48. Meng R, Gu T, Yang F, et al. Performance evaluation of the Glunovo® continuous blood glucose monitoring system in Chinese participants with diabetes: a multicenter, self-controlled trial. *Diabetes Ther* 2021; 12: 3153–3165.
49. 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: 650–657.
50. Christiansen MP, Garg SK, Brazg R, et al. Accuracy of a fourth-generation subcutaneous continuous glucose sensor. *Diabetes Technol Ther* 2017; 19: 446–456.
51. Alva S, Brazg R, Castorino K, et al. Accuracy of the third generation of a 14-day continuous glucose monitoring system. *Diabetes Ther* 2023; 14: 767–776.
52. Kim K-S, Lee S-H, Yoo WS, et al. Accuracy and safety of the 15-day CareSens Air continuous glucose monitoring system. *Diabetes Technol Ther* 2024; 26: 222–228.
53. Hanson K, Kipnes M and Tran H. Comparison of point accuracy between two widely used continuous glucose monitoring systems. *J Diabetes Sci Technol* 2024; 18: 598–607.
54. Macleod K, Katz LB and Cameron H. Capillary and venous blood glucose accuracy in blood glucose meters versus reference standards: the impact of study design on accuracy evaluations. *J Diabetes Sci Technol* 2019; 13: 546–552.
55. Kovatchev BP, Patek SD, Ortiz EA, et al. Assessing sensor accuracy for non-adjunct use of continuous glucose monitoring. *Diabetes Technol Ther* 2015; 17: 177–186.
56. Castle JR and Ward WK. Amperometric glucose sensors: sources of error and potential benefit of redundancy. *J Diabetes Sci Technol* 2010; 4: 221–225.
57. Kamath A, Mahalingam A and Brauker J. Analysis of time lags and other sources of error of the DexCom SEVEN continuous glucose monitor. *Diabetes Technol Ther* 2009; 11: 689–695.
58. King C, Anderson SM, Breton M, et al. Modeling of calibration effectiveness and blood-to-interstitial glucose dynamics as potential confounders of the accuracy of continuous glucose sensors during hyperinsulinemic clamp. *J Diabetes Sci Technol* 2007; 1: 317–322.
59. Freckmann G, Pleus S, Grady M, et al. Measures of accuracy for continuous glucose monitoring and blood glucose monitoring devices. *J Diabetes Sci Technol* 2019; 13: 575–583.
60. Eadie M and Steele RJ. Non-invasive blood glucose monitoring and data analytics. In: *Proceedings of the international conference on compute and data analysis—ICDDA '17*, Lakeland, FL, 2017, pp. 138–142. , New York, NY: Association for Computing Machinery.
61. Facchinetti A. Continuous glucose monitoring sensors: past, present and future algorithmic challenges. *Sensors (Basel)* 2016; 16: 2093.

62. Haak T, Hanaire H, Ajjan R, et al. Flash glucose-sensing technology as a replacement for blood glucose monitoring for the management of insulin-treated type 2 diabetes: a multicenter, open-label randomized controlled trial. *Diabetes Ther* 2017; 8: 55–73.
63. Bolinder J, Antuna R, Geelhoed-Duijvestijn P, et al. Novel glucose-sensing technology and hypoglycaemia in type 1 diabetes: a multicentre, non-masked, randomised controlled trial. *Lancet* 2016; 388: 2254–2263.
64. Lind M, Polonsky W, Hirsch IB, et al. Continuous glucose monitoring vs conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections. *JAMA* 2017; 317: 379.
65. Paris I, Henry C, Pirard F, et al. The new FreeStyle libre flash glucose monitoring system improves the glycaemic control in a cohort of people with type 1 diabetes followed in real-life conditions over a period of one year. *Endocrinol Diabetes Metab* 2018; 1: e00023.
66. Kröger J, Fasching P and Hanaire H. Three European retrospective real-world chart review studies to determine the effectiveness of flash glucose monitoring on HbA1c in adults with type 2 diabetes. *Diabetes Ther* 2020; 11: 279–291.
67. Pratley RE, Kanapka LG, Rickels MR, et al. Effect of continuous glucose monitoring on hypoglycemia in older adults with type 1 diabetes. *JAMA* 2020; 323: 2397–2406.
68. Ida S, Kaneko R and Murata K. Utility of real-time and retrospective continuous glucose monitoring in patients with type 2 diabetes mellitus: a meta-analysis of randomized controlled trials. *J Diabetes Res* 2019; 2019: 4684815.
69. Moström P, Ahlén E, Imberg H, et al. Adherence of self-monitoring of blood glucose in persons with type 1 diabetes in Sweden. *BMJ Open Diabetes Res Care* 2017; 5: e000342.
70. Patton SR. Adherence to glycemic monitoring in diabetes. *J Diabetes Sci Technol* 2015; 9: 668–675.
71. De Block C, Vertommen J, Manuel-y-Keenoy B, et al. Minimally-invasive and non-invasive continuous glucose monitoring systems: indications, advantages, limitations and clinical aspects. *Curr Diabetes Rev* 2008; 4: 159–168.
72. Kim J, Campbell AS and Wang J. Wearable non-invasive epidermal glucose sensors: a review. *Talanta* 2018; 177: 163–170.
73. Price C, Ditton G, Russell GB, et al. Reliability of inpatient CGM: comparison to standard of care. *J Diabetes Sci Technol* 2023; 17: 329–335.
74. Clubbs Coldron B, Coates V, Khamis A, et al. Use of continuous glucose monitoring in non-ICU hospital settings for people with diabetes: a scoping review of emerging benefits and issues. *J Diabetes Sci Technol* 2023; 17: 467–473.
75. Wang M, Singh LG and Spanakis EK. Advancing the use of CGM devices in a non-ICU setting. *J Diabetes Sci Technol* 2019; 13: 674–681.