Journal of Diabetes Science and Technology
2020, Vol. 14(1) 135–150

DOI: 10.1177/1932296819855670
journals.sagepub.com/home/dst
SAGE

Benefits and Limitations of MARD as a Performance Parameter for Continuous Glucose Monitoring in the Interstitial Space

Lutz Heinemann, PhD[1], Michael Schoemaker, PhD[2], Günther Schmelzeisen-Redecker, PhD[2], Rolf Hinzmann, MD, PhD[2], Adham Kassab, MSc[3], Guido Freckmann, MD[4], Florian Reiterer, PhD[5], and Luigi Del Re, PhD[5]

Abstract

High-quality performance of medical devices for glucose monitoring is important for a safe and efficient usage of this diagnostic option by patients with diabetes. The mean absolute relative difference (MARD) parameter is used most often to characterize the measurement performance of systems for continuous glucose monitoring (CGM). Calculation of this parameter is relatively easy and comparison of the MARD numbers between different CGM systems appears to be straightforward on the first glance. However, a closer look reveals that a number of complex aspects make interpretation of the MARD numbers provided by the manufacturer for their CGM systems difficult. In this review, these aspects are discussed and considerations are made for a systematic and appropriate evaluation of the MARD in clinical trials. The MARD should not be used as the sole parameter to characterize CGM systems, especially when it comes to nonadjunctive usage of such systems.

Keywords

continuous glucose monitoring, diabetes therapy, quality of measurement, glucose measurement, blood glucose

Many patients with diabetes routinely use systems for continuous glucose monitoring (CGM) as the diagnostic cornerstone of their diabetes treatment and the number is assumed to increase massively in the next years. CGM systems have seen considerable improvements in their performance after their market introduction some 15 years ago. Their analytical performance, that is, the accuracy with which CGM systems can measure glucose concentrations in interstitial fluid (ISF) in the subcutaneous adipose tissue was also improved. Such improvements have tremendous importance for clinical utility of CGM systems, mainly in view of their use with a nonadjunctive claim, that is, that therapeutic decisions can be based on the glucose values presented without exposing patients to too high risks. CGM data are also used in other clinical applications like bolus calculators (for patients on multiple daily insulin injections), in insulin pumps (with or without automatic changes in insulin infusion rates depending on the current glucose measurement results), and—last but definitively not least—are indispensable in automated insulin delivery (AID) systems.

Time in range(s) (TIR) has been suggested by several working groups as a new "biomarker" to complement HbA1c for glycemic control. There are good arguments that TIR has advantages compared to HbA1c since

(1) it responds faster to treatment changes
(2) it reflects as well glycemic variability, hypo- and hyperglycemia
(3) it is not affected by physiological and pathological factors that affect the HbA1c concentration

In order to establish TIR successfully in clinical practice and in order to make it a useful endpoint in clinical studies it is important that it can be measured accurately. For "time"

[1]Science & Co, Neuss, Germany
[2]Roche Diabetes Care, Mannheim, Germany
[3]mySugr GmbH, Wien, Austria
[4]Institut für Diabetes-Technologie Forschungs- und Entwicklungsgesellschaft an der Universität Ulm, Ulm, Germany
[5]Johannes Kepler Universität, Linz, Austria

Corresponding Author:
Lutz Heinemann, PhD, Science & Co, Geulenstr 36, 41462 Neuss, Germany.
Email: l.heinemann@science-co.com

this is no problem; however, the accuracy of CGM systems strongly affects the ranges. If results differ between different CGM systems, TIR calculated from these results will do so as well. There is a clear need to better assess the analytical accuracy of CGM systems than this can be done by means of the MARD and to establish metrological traceability for CGM measurements. For systems used for self-monitoring of blood glucose (SMBG), an ISO standard was established some decades ago to characterize their analytical performance. However, no such standard was established for CGM systems until now, even if attempts have been made.[1] The parameter most often used for description of the analytical performance of CGM systems is the "mean (or median) absolute relative difference" (MARD). Reasons why this parameter is widely used is the relatively ease with which this parameter can be calculated and that a single number if presented that appear to enable a straightforward interpretation of the MARD of a given CGM system and allow comparison of the performance of different CGM systems.

In this review, it will be discussed that in reality a large number of aspects makes the MARD difficult to interpret.

MARD—Background and Properties

CGM systems measure glucose in the interstitial fluid in the subcutaneous tissue while SMBG measures glucose in capillary blood. The assumption is, that glucose levels in both compartments (matrices) are similar; however, this holds true for steady state conditions only, which is not the case most of the day of patients with diabetes. To assess the accuracy of a measurement one has to make sure the "measurand" is identical, that is, (1) same matrix, (2) same substance, (3) same units are used. Otherwise a comparison is metrologically not allowed. For this reason, all manufacturers use algorithms that predict BG values (typically capillary blood values and not venous blood values) from CGM values. This is why CGM values are compared to SMBG values.

The MARD is computed using temporally matched glucose data from CGM systems and comparison glucose measurements (most often obtained by capillary blood glucose (BG) measurements) of all subjects of a clinical study (Figure 1). It is important to note that the MARD is a measurement of the performance of the system (= sensor + algorithm) rather than the sensor element alone. The impact of the algorithm should not be underestimated—its design allows trading-off accuracy against "drop out time" of the CGM system. For redundant sensor configurations, a "joint" MARD can be obtained as the processed output of the multiple individual sensors with some weight, for example, according to the "health" of the sensor.

Reported as a percentage, MARD is the average of the absolute difference between these values. The less the MARD is, the closer are the CGM readings to the comparison values. Typically a CGM system with a MARD <10% is regarded to have good analytical performance. The MARD

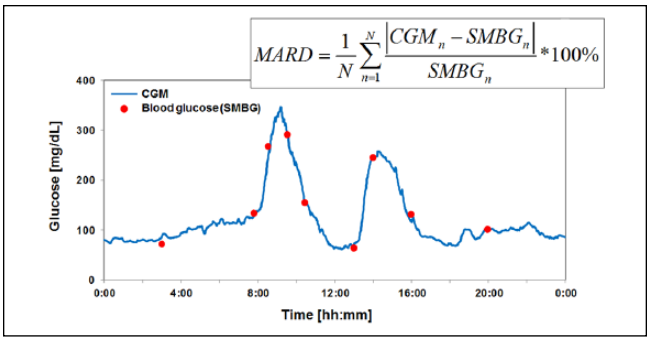

Figure 1. Glucose values continuously monitored (CGM) over 24 hours and corresponding BG values measured by a comparison method in certain time intervals. Based on the absolute differences between the CGM and BG glucose (in this case SMBG, but can also be a different comparison method) the MARD is calculated according to the formula given.

Table 1. Factors That Influence the Assessment of the MARD.

CGM system-inherent factors (intrinsic performance of the system)
• Calibration
• Performance of the CGM sensor over time
• Sensor to sensor variation
• Algorithms and smoothing filters implemented in the CGM systems
Not CGM system-inherent factors
• Insertion factors (competency of sensor insertion, body site, movement)
• Physiological time delay between CGM and BG measurements
• Range and distribution of the paired glucose values
• Rate of change in glycemia
• Study design/ study population
• Implementation of the study design (controlled vs uncontrolled environment, time of day, day 1 vs day 2-X, dynamic variation in glucose concentration, number of measurements, synchronization of CGM, and comparison method)
• Direction of the deviations from the comparison method

is a statistical approach used in other respects as well; however, it is not used often to characterize the performance of systems for SMBG as it doesn't distinguish between precision and bias (and is therefore also not mentioned in the respective ISO standard) and in other areas of diabetes research.[2] It is worth to note that the MARD does not differentiate between positive and negative errors or between systematic and random errors. In other words, the MARD is influenced by a number of factors (Table 1) and has a number of advantages and disadvantages (Table 2).

In the last years practically in each publication about CGM systems, a MARD value was reported. However, an attempt to find publications about MARD and its properties itself—with a focus on diabetes—was not very successful: A literature search in PubMed for this term revealed a limited number of such publications. Nevertheless, some of our own publications[3-8] and others[9,10] do discuss this topic. It is of

Table 2. List of Advantages and Disadvantages of the MARD.

Advantages
- Provide information about analytical performance of CGM systems in one number
- Widely used
- Perceived as a parameter enabling comparison of the performance of different CGM systems
Disadvantages
- Only a small portion of CGM data is used for calculation of MARD
- Addition information provided by CGM systems like trend and rate-of change are not taken into account
- The same holds true for frequency and relevance of artifacts (eg, signal dropouts) that most often will not be recognized by relatively seldom performed comparison measurements
- Although MARD is influenced by both accuracy metrics, bias and precision of CGM measurements, it does not allow to distinguish between the two
- MARD does not differentiate between a positive or negative bias to the comparison measurements
- Overall MARD does not provide information during the time between the sporadic comparison measurements
- MARD does not provide specific information during time periods with dynamic changes in glycemia or during episodes of hypo- or hyperglycemia
- MARD does not provide information about transient large sensor inaccuracies (induced by a calibration issue or movement artifacts)
- MARD does not reflect durability of the sensor

interest to note that different studies with the same CGM system provide impressively different MARD results (Table 3 and Figure 2). For example, the MARD of the Dexcom G4 has been reported as everything between 11% and 21%, without any regularity over time which could be explained as an improvement of the sensor (eg, the first batch being worse than subsequent versions). In the following different aspects will be discussed that are of relevance for interpreting the MARD and that explain why such differences in MARD have to be expected.

Median or Mean

Usage of the median instead of the mean absolute relative difference (MARD) has pro and cons: outliers have reduced impact on the median in comparison to the mean value; however, conversely, if a data set has a several outliers, these are somewhat hidden by using the median. Usually the median ARD is lower than the mean ARD, the median ARD is nearly 0.8 multiplied by the MARD. This relationship been observed both empirically for multiple data sets and has been shown theoretically.[9] Outliers have essentially no influence on the median ARD and only a small influence on MARD. From a clinical point of view or better from a patient point of view, outliers are of relevance when the patient is adapting his therapy on such a value. In publications, it is often not clearly stated if the given MARD represents the median or mean.

Glucose Range

Due to differences in the analytical performance of the CGM system and the comparison measurement system when it comes to measurement of glucose in the hypo-, eu-, and hyperglycemic range (plus the fact that the CGM system measures glucose in ISF and the comparison system in blood), the MARD of a given CGM system might differ between these ranges, that is, the MARD can be worse in the hypoglycemic range in comparison to the euglycemic range.[9] Thus, not only an overall MARD should be reported, but also stratified MARD values for different glucose ranges. As an alternative way to account for this effect, it has recently been proposed to normalize MARD based on some comparison distribution for paired glucose measurements in order to make values from different clinical trials more comparable.[8]

Precision of CGM Systems

A key advantage of CGM is the high frequency of measurement. However, as MARD requires using temporarily matched comparison values, and comparison values are not available with the same frequency, many CGM values are not used. An alternative metric, the precision of CGM systems, overcomes this limitation by using the data obtained from two CGM systems (of the same make and model!) worn in parallel by the same patient; this enables calculation of the "precision absolute relative difference" (PARD; see below).[5,6,19] An additional advantage of the PARD—besides making use of all CGM data available—is that it does not measure glucose in two different compartments (= all physiological differences [see below] are not of relevance).

In general, it is appropriate to assume that both CGM systems have more or less identical errors; however, also different CGM systems can be used (as an attempt to have a continuous comparison measurement signal instead of the relatively sparse BG measurements) (see below), but of course we may not assume that the errors will be the same.

Timing of Comparison Measurements and Quality of such Measurements

For the calculation of the MARD CGM values are brought in relation to the "true" glucose value measured in blood samples. The underlying assumption of this approach is that the " comparison measurements" (ie, those produced by the comparison method) are inherently error free—which in reality is not true, as also this comparison method has a more or less large measurement error. Thus, the MARD does not only reflect the analytical performance of the CGM system under test, but also that of the comparison measurement. In other words, the calculation of the MARD includes the influence of both bias (= accuracy) and precision. As an example, consider two perfectly accurate but different CGM systems (when using one CGM system but measure with

Table 3. Listing of Clinical Trials With Different CGM Systems Summarizing Key Study Information and the MARD Obtained.

Reference	Study setting	Number of subjects	T1/T2	Study duration	Comparison method	Calibration system	Number of calibrations per day	Overall MARD	MARD during hypos	Comments
Dexcom seven plus										
Garg[11]	Home use, with 3 in-clinic visits	14	T1	15 days (5 days/sensor)	Venous measurements (YSI); SMBG	BG meter	2 h after sensor insertion; every 12 h thereafter	vs YSI: 16.8%, vs BG: 16.2%	Hypo (<80 mg/dL), YSI: 21.5%	
Luijf[12]	Both (1 day in clinic and rest is home use)	12	T1	7 days	In clinic: plasma glucose (YSI); Home use: SMBG (Aviva Nano)	Accu-Chek Aviva Nano	Twice per day	In clinic: 19.2%; Home use: 16.8%	In clinic: 23.9%; Home use: 41.6	
Freckmann[13]	In clinic	12	T1	9 days (7 days for Dexcom Seven Plus)	Capillary BG and venous BG (YSI)	Navigator BG meter	2 h after sensor insertion and then every 12 h	16.7%	31.7%	
Facchinetti[14]	In clinic	24 (12 from each database)	T1	7 days	BG samples measurements (YSI)	Not reported	According to the manufacturer's instructions	15.1% (15.6% study 1 & 14.7% study 2)		Study looked into two databases from 2 studies
Damiano[15]	In clinic	6	T1	2 days	Venous plasma glucose (GlucoScout and YSI)	Not reported	4 calibrations with venous blood	16.5%		
Christiansen[16]	In clinic	53	T1 (81%) & T2 (19%)	7 days	Venous measurements (YSI)	OneTouch Ultra 2 meter	Twice daily	16.0%	27.0%	
Facchinetti[17]	In clinic	9	T1	1 day	BG samples measurements (YSI)	Not reported	Twice per day	14.1%		Median MARD
Dexcom G4 platinum										
Luijf[18]	Clinic (from 8 am until noon) and home use	20	T1	Clinic (1 day) and Home use (7 days)	Venous measurements (YSI)	Abbott Freestyle BG meter	According to manufacturer's recommendation	Clinic: 20.5%, Home use: 16.5%		Dexcom G4 version A sensor (G4A)
Christiansen[16]	In clinic	72	T1 (83%) & T2 (17%)	7 days	Venous measurements (YSI)	OneTouch Ultra 2 meter	Twice daily	13.0%	18.0%	
Pleus[19]	In clinic	10	T1	7 days	Capillary BG measurements	Accu-Chek Aviva	2 h after sensor insertion and then every 12 h	11.0%	13.7%	
Garcia[20]	In clinic	72	Not reported	7 days	Venous measurements (YSI)	Not reported	Twice daily	13.2%		
Damiano[21]	In clinic	24 (12 adults & 12 children)	T1	2 days	Venous measurements (GlucoScout)	GlucoScout meter	One calibration before the start of the experiment, and during the experiment	10.8%		
Matuleviciene[22]	Home use	38	T1	4-6 days	Capillary and venous BG measurements	HemoCue	every 12 h	13.9%	20.0%	
Kropff[23]	In clinic phase (6 h) and home use phase	24	T1	6 days	Capillary BG measurements & 6 h on the 3rd day compared to venous blood (YSI)	Accu-Chek Aviva	Twice a day after initial calibration	Clinic: 13.6%, Home: 12.2%	Clinic: 17.6%, Home: 21.2%	
Facchinetti[17]	In clinic	36	T1 (32) & T2 (4)	7 days	BG samples measurements (YSI)	Not reported	Twice daily	11.2%		Median MARD
Pleus[24]	In clinic	10	T1	7 days	Capillary BG measurements	Accu-Chek Aviva	2 h after sensor insertion and then every 12 h	10.9%		

(continued)

Table 3. (continued)

Reference	Study setting	Number of subjects	T1/T2	Study duration	Comparison method	Calibration system	Number of calibrations per day	Overall MARD	MARD during hypos	Comments
van Beers[25]	In clinic (3 sessions) and the rest Home use	72	T1 & T2	Not reported	YSI for in clinic and SMBG for home use	Not reported	YSI for in clinic and SMBG for home use	13%		
Laffel[26]	Both	176	T1 or T2 (only 1 patient)	7 days	BG measurements or venous measurements (YSI)	LifeScan OneTouch Verio IQ	Twice daily	17% vs YSI. & 15% vs BG	18% vs YSI. & 21% vs BG	Pediatric subjects, two CGMs with one receiver and masked data.
Andelin[27]	In clinic	41	T1	6 days	Capillary and venous blood samples (HemoCue)	HemoCue	Capillary glucose	Capillary: 11.7%; Venous: 13.7	Capillary: 16.6%; Venous: 31.8%	
Bonora[28]	Home use	8	T1	14 days	BG capillary blood	Not reported	At least twice daily	vs SMBG: 15.7%, vs FSL: > Day 1-10: 17.7%; > Day 11-14: 22.2%	FSL vs DG4P: 18.7%	Goal: Determine agreement between FSL and DG4P
Aberer[29]	In clinic	12	T1	12 h	CGM and analyzing venous plasma glucose values every 5 min (Super GL Glucose Analyzer)	Capillary BG	every 12 h	16.8%	23.8%	
Dexcom G5 (or G4 Platinum sensor + Software 505)										
Bailey[30]	Both	51	T1 (86%) & T2 (14%)	7 days	During clinic session: YSI, during home use: BG	Bayer Contour Next USB	Twice daily	vs YSI: 9.0% vs BG: 11.2%		
Laffel[26]	Both	79	T1	7 days	BG measurements or venous measurements (YSI)	Bayer Contour Next USB	Twice daily	vs YSI 10% & vs BG 13%	vs YSI 17% & vs BG 19%	Pediatric subjects, single unmasked sensor
Freckmann[31]	Home use with 3×48 hours in-clinic visits	20	T1	14 days (7 days/sensor)	SMBG (14 days) / SMBG + venous BG (Hexokinase (HK)) during 3x 6 hours induced glucose excursions	FreeStyle Freedom Lite	2 h after sensor insertion and then every 12 h	14 days: 12.4%; Glucose excursions: vs HK: 16.2% vs SMBG: 14.7%		
Dexcom G6										
Wadwa[32]	In clinic phases for CGM assessment, otherwise patients were at home	262	T1 (99%) & T2 on insulin (1%)	10 days	YSI	None (glucose values were computed retrospectively based rely on measured sensor currents)	None	10.0%		

(continued)

Table 3. (continued)

Reference	Study setting	Number of subjects	T1/T2	Study duration	Comparison method	Calibration system	Number of calibrations per day	Overall MARD	MARD during hypos	Comments
Medtronic SofSensor										
Welsh[33]	Not reported	71 adults (19-72) & 61 children (7-17)	T1	3 days	Not reported	Paradigm Link BG monitor	Not reported	Adults: 9.9%; Children: 10.1%		
O'Neal[34]	Home use	10	T1	3 days	(YSI)	Contour Link Meter, Bayer	Not reported	18.9%		Study is only looking into nocturnal hypoglycemia, no overall MARD
Christiansen[16]	In clinic	72	T1	4 nights (from 18:00 until the next morning)	venous BG (Roche Hitachi 912 chemistry analyzer)	Not reported	Twice every night		45.0%	Using the iPro algorithm
Medtronic Enlite I										
Kovatchev[35]	In clinic	14	T1	Not reported	YSI	YSI BG analyzer	According to manufacturer's recommendation	15.2%	16.1%	
DirectNet Study Group[36]	Most at home and 18 h in clinic	30	T1	7 days	BG at home and Laboratory serum glucose in clinic	One Touch® Ultra® Meter (at home)	Every 12 h at least	14% (in clinic); 17% (at home)		Children and youth (4-17 years); Median MARD
Keenan[37]	In clinic	98	T1 (79) & T2 (18)	7 days	Plasma glucose measurements (YSI)	Not reported	Not reported	G: 13.1% & 14.6%; V: 12.9% & 14.7%		Data were processed using the calib. Algo. of Guardian (G) and the Veo real time (V)
Calhoun[38]	In clinic	54	T1	69 nights	19% YSI analyzer, 66% GlucoScout monitor, and 15%; HemoCue analyzer	Not reported	Manufacturer-recommended intervals	15%		Median MARD
Luijf[18]	In clinic phase (from 8 am until noon) and a home use phase	20	T1	Clinic phase (1 day) and Home use phase (6 days)	(YSI glucose analyzer)	Abbott Freestyle BG meter	According to manufacturer's recommendation	Clinic: 16.4%; Home use: 18.9%	Clinic: 21.5%	
Damiano[15]	In clinic	6	T1	2 days	Venous BG measurements (GlucoScout and YSI)	No reported	Average of 4.7 calibrations with venous blood	18.0%		
Freckmann[39]	In clinic	12	T1	9 days (6 days for Enlite)	capillary BG measurements and venous BG (YSI)	Navigator BG meter	1, 2, 10, 24, 72 hours after sensor insertion	16.7%	34.3%	
Mahmoudi[40]	In clinic	10	T1	Not reported	Plasma glucose (HemoCue glucose analyzer)	Not reported	At least three times daily	32.1%	62.8%	Glucose data were restricted to nighttime measurements
Damiano[21]	In clinic	24 (12 adults & 12 children)	T1	2 days	Venous BG measurements (GlucoScout)	GlucoScout meter	One calibration before the start of the experiment, during the experiments	17.9%		

(continued)

Table 3. (continued)

Reference	Study setting	Number of subjects	T1/T2	Study duration	Comparison method	Calibration system	Number of calibrations per day	Overall MARD	MARD during hypos	Comments
Bailey[41]	In clinic (3 sessions) and the rest Home use	90	T1 & T2	6 days	YSI	Bayer Contour NEXT LINK meter	Different frequencies	1.2 calibrations daily: 14.7%; 2.8 calib. daily: 13.6%		
Matuleviciene[22]	Home use	38	T1	4-6 days	Capillary and Venous BG measurements	HemoCue	According to manufacturer's recommendation	17.9%	34.7%	
Kropff[23]	In clinic phase (6 h) and a home use phase	26	T1	6 days	Capillary BG fingersticks & 6 h on the 3rd day compared to venous blood (YSI)	Accu-Chek Aviva	Twice a day after initial calibration	In clinic phase: 16.6%, Home phase: 19.9%	In clinic phase: 24.6%, Home phase: 36.5%	
Nørgaard[42]	Home use	45	T1	15 days (3 days/device)	BG capillary blood	Not reported	Mean number of 4.7 per day	15.5%		
Medtronic Guardian Sensor 3										
Aberer[29]	In clinic	12	T1	12 hours	CGM and analyzing venous BG values every 5 min (Super GL Glucose Analyzer)	Capillary BG calibration	At least twice a day	21.4%	26.9%	Guardian Sensor 3 as part of the Medtronic MiniMed 640G
Christiansen[43]	In clinic (3 sessions / 12-14h) and the rest Home use	88	T1 (62) & T2(26)	7 days	YSI FST; using 2300 STAT Plus	Home: not reported; During FST: sensors were calibrated based on prompts from the MiniMed 640G	FST: 40-120 min after sensor insertion, 6 h and 12 h after first calibration	Minimum calibrations: 10.6%; one additional calibration: 9.6%		Guardian Sensor 3 as part of the Medtronic MiniMed 640G
Abbott Navigator 1										
Weinstein[44]	In clinic	58	T1	5 days	Venous YSI measurements	Integrated FreeStyle meter	Initially at 10 h post insertion; 12, 24, and 72h after insertion	12.8%	19.8%	
Kovatchev[35]	In clinic	14	T1	Not reported	YSI	YSI r	According to manufacturer's recommendation	15.3%	10.3%	
Garg[11]	Home use with 3 in-clinic visits	14	T1	15 days (5 days/sensor)	Venous YSI; SMBG	FreeStyle integrated BG meter	Initially at 10 h post insertion; 12, 24, and 72h after insertion	vs YSI:16.1%; vs BG: 15.9%	hypo (<80 mg/dL) vs YSI: 29.8%	
Luijf[18]	In clinic phase (from 8 am until noon) and a home use phase	20	T1	Clinic (1 day) and Home use (5 days)	(YSI)	Abbott Freestyle BG meter	According to manufacturer's recommendation	Clinic: 16.5%; Home: 14.5%	Clinic: 17.4%	
Damiano[15]	In clinic	6	T1	2 days	Venous BG measurements (GlucoScout and YSI)	No reported	One calibration with venous blood	11.8%		
Freckmann[13]	In clinic	12	T1	9 days (5 days use for Navigator)	Capillary BG and venous BG (YSI)	Navigator BG meters	1, 2, 10, 24, 72 hours after sensor insertion	12.6%	23.3%	
Damiano[21]	In clinic	24 (12 adults & 12 children)	T1	2 days	Venous BG (GlucoScout)	GlucoScout meter	Two calibrations before the start of the experiment, and during the experiment an average of 6.4	12.3%		

(continued)

Table 3. (continued)

Reference	Study setting	Number of subjects	T1/T2	Study duration	Comparison method	Calibration system	Number of calibrations per day	Overall MARD	MARD during hypos	Comments
Abbott Navigator II										
McGarraugh[45]	In clinic	47	T1	10 h	Venous BG vs YSI	Not reported	Not reported	14.5%	17.5%	
Leelarathna[46]	AID and conventional treatment or two AID visits	32 (12 pregnant)	T1	Study visits: 18 h to 36 h	(YSI)	Integrated Freestyle BG meter	According to manufacturer's recommendation	13.9%	Only median ARD: 21.0%	Analyzed data from five AID studies
Thabit[47]	Home use	57	T1	2 periods in the overnight studies, and 1 week for the day-and-night study	Capillary BG fingersticks	Integrated Freestyle BG meter	According to manufacturer's recommendation	14.2%	30.6%	Accuracy was assessed using data from 3 studies
Abbott Freestyle Libre										
Bailey[48]	Home use with 3 in-clinic visits	72	T1 & T2	14 days	Capillary BG / venous BG (YSI)	Factory-calibrated	No calibration during 14 days of wear	vs BG: 11.4%; clinic alone with BG and YSI: 12.1% and 12.0%		
Bonora[28]	Home use	8	T1	14 days	BG capillary blood	Factory-calibrated	No calibration during 14 days of wear	vs BG: 15.4%, vs DG4P: >Day1-10: 17.7%; >Day 11-14: 22.2%	FSL vs DG4P: 18.7%	Determine agreement between FSL and DG4P, >3 sensor scans per day at <8 h intervals
Aberer[29]	In clinic	12	T1	12 h	CGM and BG values every 5 min (Super GL Glucose Analyzer)	Factory-calibrated	Not reported—max. wear time: 14 days	13.2%	14.6%	Sensor on the back of each upper arm (two sensors total); no significant difference
Freckmann[31]	Home use with 3×48 hours in-clinic visits	20	T1	14 days (7 days/sensor)	SMBG (14 days) / SMBG + venous BG (Hexokinase (HK)) during 3x 6 hours induced glucose excursions	Factory-calibrated	No calibration during 14 days of wear	14 days: 13.0%; Glucose excursions: vs HK: 16.8%; vs SMBG: 15.7%		
Senseonics Eversense										
Kropff[49]	Home use with 11 clinic visits	71	T1 (66) & T2 (5)	180 days	Venous measurements (YSI)	Accu-Chek Aviva	Twice daily	11.6%	21.7%	
Christiansen[50]	Home use with 7 clinic visits	90	T1 (61) & T2 (29)	90 days	Home: BG (Contour next) Clinic: YSI	Contour Next USB	Twice daily	8.8%	40-54 mg/dL: 10.7%, 55-70 mg/dL: 9.0%	

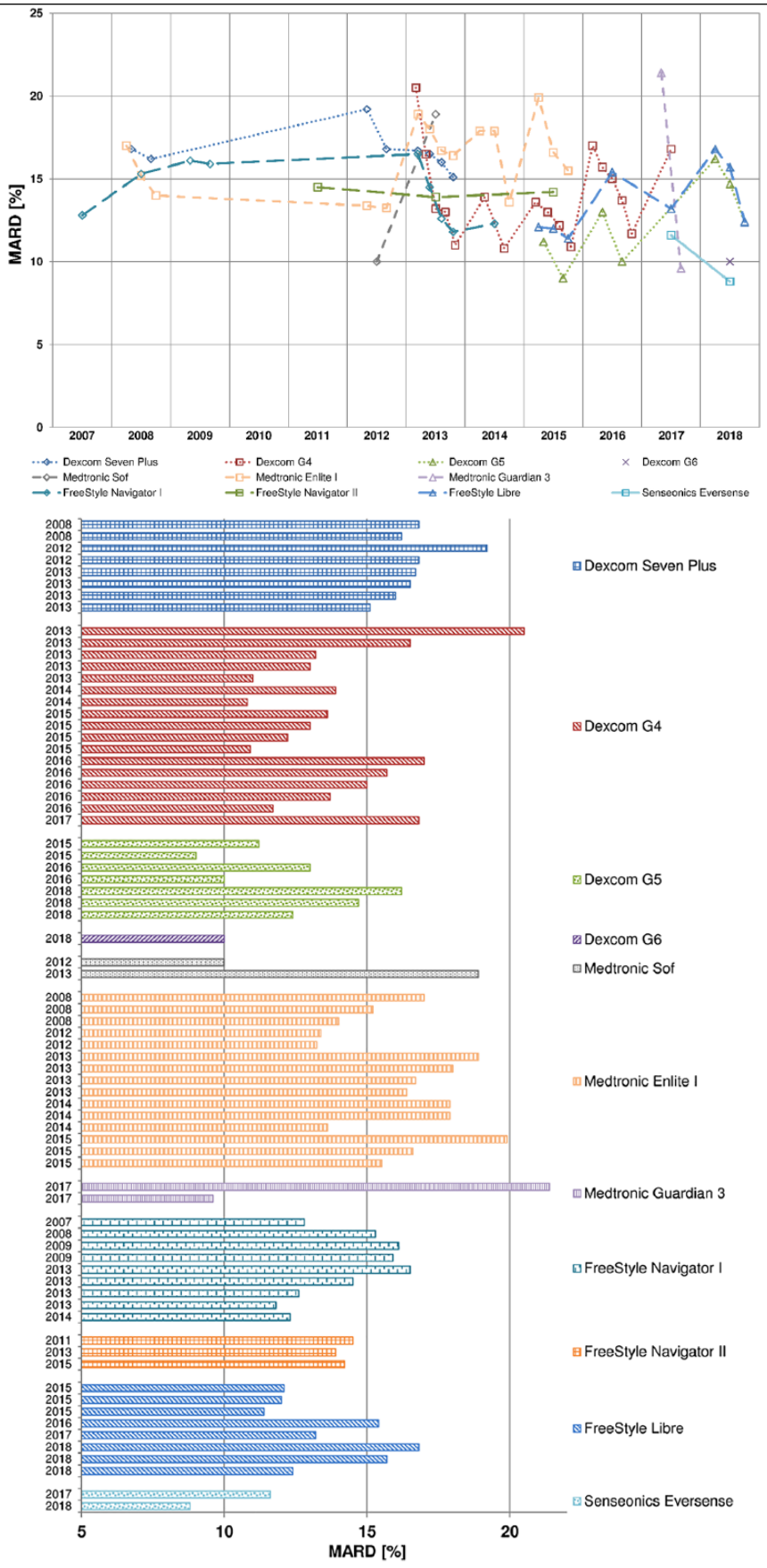

Figure 2. *(continued)*

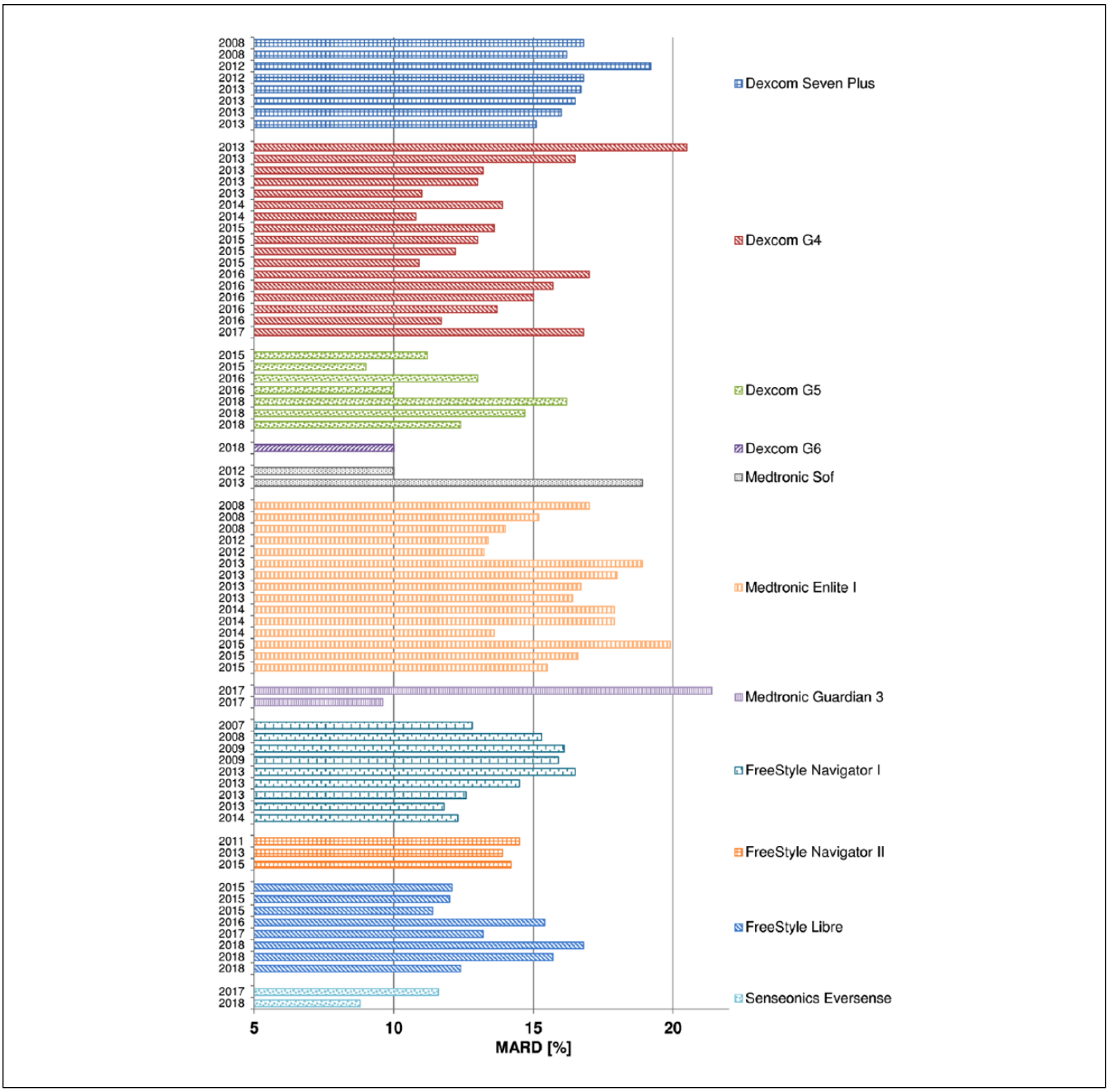

Figure 2. Graphical representation of the MARD values obtained in the different studies listed in Table 3. (a) MARD values separated for the different CGM systems and year the respective studies were published. (b) MARD values sorted by CGM manufacturer / CGM system and year the respective studies were published.

two devices in the same subjects, the PARD can be calculated; see below) with an expected bias of zero across the entire range of measurement, but with some random measurement errors. If these two systems are tested against each other and the MARD is computed from the paired measurements, this MARD will not be zero—it will reflect the presence of the measurement errors. One would need to have both "perfect accuracy" (no bias) and "perfect precision" (no measurement errors) in order to obtain a MARD of zero. It might also be, that two quite different CGM systems have the same MARD, but are quite different with respect to bias and precision: one has a good precision, but a large bias and the other has a small bias, but a bad precision. What is clinically more relevant?

More precisely, some additional aspects have to be considered:

1. The error distribution of the test method (= CGM system) and of the comparison method should be similar and the magnitude of bias and random error of the comparison method should be lower. If the comparison method is a good laboratory method, then its error might be $<\pm 2\%$; however, if BG systems are used as comparison systems, as it is frequently done, their errors are in the range of $\pm 5\%$ to $\pm 15\%$. As most recent CGM systems have a low measurement error, their analytical performance comes close to that of at least some comparison methods, but for reasons of principle can never reach this. Especially at low glucose levels, the random error of the comparison method may be particularly high and, therefore, important to consider.
2. The number of comparison measurements has a massive impact on the MARD. Therefore, results obtained in clinical studies under highly controlled conditions and frequent comparison measurements differ from that obtained under real-world conditions (with more sparse comparison measurements and most often usage of a comparison method with a lower measurement quality).
3. As already discussed, the distribution of the comparison measurements in the glucose value range is of high importance.
4. The MARD does not consider the time periods between comparison measurements, that is, there might be a higher number of parallel measurements during daytime, but only a limited number during the night. It is not clear immediately if the performance of CGM systems is equally good during both times, especially because during night time there may be additional mechanical stresses as the patient may induce a compression artefact when lying on the glucose sensor.

Dynamic Changes in Glycemia

Patients with diabetes exhibit more or less pronounced glucose swings in daily life, that is, glycemia changes over time with different rates. Glucose variability make patients also prone to develop acute glycemic deteriorations into clinical relevant ranges outside the physiological range; however, there are considerable differences in the risk of patients (depending on many other factors) to become hypo- or hyperglycemic.[5,19,51]

It is important to acknowledge that the measurements for MARD calculation are performed in two different compartments. When glucose levels are more or less constant (rate of glucose change <1 mg/dl/min), the glucose levels in BG and ISF are practically identical. However, in case of swings in glycemia (= rate of change >1 mg/dl/min) physiological differences in glucose levels between both compartments show up. In other words, a rapid increase in glycemia after eating a meal with rapidly absorbable carbohydrates does not result in an immediate increase in CGM recordings; a decrease in peripheral glucose levels during exercise is not directly reflected in a decline in BG levels, a not removable (albeit predictable) time delay shows up.[3] Such differences can lead to relevant differences in insulin dose selection depending on which glucose signal the selection is based on.[52]

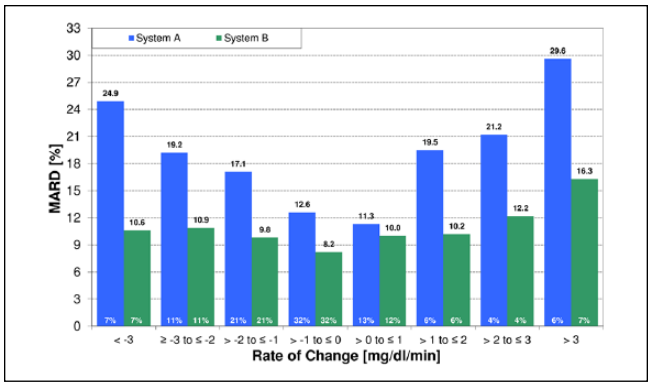

Figure 3. MARD during glucose swings (= periods of rapid glucose changes) for two different CGM systems as a function of the rate of change.[19]

The extent of differences in numbers does not only depend on the rate of change in glucose levels and the CGM system studied (Figures 2 and 3), it can also depend on the period of the day and different body sites. The knowledge about how constant this time delay is and to which extent it differs inter- and intraindividually is limited.[3,7] An additional "physical" time delay is introduced by the measurement technique and data filtering plus analysis; however, this delay is in the range of seconds or minutes with most CGM systems. The algorithms implemented into the CGM systems try to compensate for this delay. The reported CGM values are an estimated/predicted BG value. This is done since (1) patients are used to interpreting BG values and (2) BG values are used for calibrating and (3) verifying CGM values.

The impact of the rate of change of glycemia on MARD values has two reasons:

1. The time delay seen by a CGM system expresses itself as a "wrong" concentration measurement, which is in fact not true.
2. The additional physical time delay of a given CGM system (with considerable differences between them due to different diffusion times, algorithms used to calculate the glucose values from the current measured by the glucose sensor by using a calibration algorithm while trying to reduce noise at the same time, etc) can make this difference between BG and ISF glucose larger or smaller. In the example shown in Figure 3 the shorter physical time delay of one CGM system leads to a lower MARD than that of the other CGM system with a longer time delay. At the same time, the overall MARD of both CGM systems was relatively similar (9% vs 11%).

Currently therapeutic decision-making is based on SMBG-related thinking, that is, the assumption that a given BG spot measurement provides "reliable" insights into the glucose changes over time. Such spot measurements are made with high accuracy; however, they provide only snap shots of changes over time in comparison to CGM systems which provide a complete picture. The lower accuracy of glucose concentration measurements with CGM systems is compensated by the additional information (trend arrows, alarms) provided. Nevertheless, also these CGM measurements are affected with errors. The MARD only considers the error in glucose concentration measurement and does not take these additional CGM information into account; however, when therapeutic decisions are based on CGM data, they can be taken into account.[52]

Compensation of the time delay during analysis of clinical trial data leads to lower MARD values. Manufacturers of CGM systems make attempts to do so as well during daily life usage of CGM systems by using "smart" sensors; such CGM systems use algorithms to improve their performance.[14]

Differences between glucose values measured at the same point in time with a BG system and a CGM system can be disturbing for patients (and diabetologists) if they are not aware of the physiological background of these.[52] They might lose confidence in the technology. This can be even worse if alarms are generated by the CGM systems that are not confirmed by BG measurements. Such issues have to be addressed in CGM teaching and training programs.[53]

Performance of an Individual Glucose Sensor

The MARD does not allow making an accurate statement about the performance of given glucose sensors of one type of CGM system. Relatively high numbers of the standard deviation (as a reflection of the differences in MARD between sensors) often given with the calculated mean (or median) MARD indicate that considerable differences between the performances of glucose sensors show up in reality. However, MARD does not distinguish whether they are due to the analytical performance of the sensors themselves or if they reflect physiological factors (and movement artifacts) at the insertion site of the sensors (see below).

In other words, the MARD provides a number that reflects the total performance of a given CGM system in a clinical study, but is not as accurate when it is used to characterize the specific sensor alone. This in turn has relevance for the clinical usage of CGM systems; the therapeutic decision in a given moment is based on the glucose value provided by the sensor, but the error of this value will not be the same for every patient and sensor of the same make.

Calibration

The performance of a glucose sensor is potentially massively influenced by the calibration procedure; this can induce a systematic deviation with respect to the measurement results obtained with the comparison method. CGM systems that require calibration need BG measurements for this purpose in regular intervals after the glucose sensor was inserted ("in vivo" calibration). Two broadly used systems (FreeStyle Libre and Dexcom G6) are calibrated during the manufacturing process. Patients don't need to perform any BG measurements when using such CGM systems; however, they have an option to do so in case of the G6. Prerequisite is a homogenous manufacturing of the glucose sensors over time and the assumption that the tissue conditions in the subcutaneous space within a patient and between patients are also "constant"; that is, the ratio between the in vitro and the in vivo sensitivity of a CGM sensor needs to be constant over time both within-patients and between patients

Optimal calibration of CGM systems requires that BG measurements are performed adequately (ie, avoiding user errors) with BG systems with a good quality. Performance of a "quality check" by BG measurements once per day helps ensure that the measurement of a given CGM system is not too far removed from the BG values; even for current CGM systems where calibration is not mandatory. In clinical studies with CGM systems, the same BG system should be used for performing calibration measurements by different patients to avoid additional sources of variability.

Consistency of MARD Over Time of CGM Usage

After insertion of the glucose sensor through the skin into the subcutaneous space it usually takes some hours until reliable measurement results are possible ("run-in phase").[19] The MARD in this period of time is high and declines in thereafter, that is, the analytical performance is better after this run-in phase for a number of days before it starts to worsen. Depending on which data are used for MARD calculation (including those of the run-in phase or not) and those of the days before the decline in performance starts (usually the recommended duration of use) the MARD might differ. Not only a summary MARD should be provided, but also calculations for all usage days (ie, MARD per day). Such a stratification over time provides instructive information about the changes in analytical performance of different CGM systems.[54]

The measurement performance of a given glucose sensor is also influenced by the conditions it is exposed to in the subcutaneous space. If the sensor tip is moved around all the time by movements of the patient (including "micro movements" that are not visible) or pressure is applied on the skin area around the sensor insertion site (eg, during sleep or sitting), this can affect tissue physiology such as exchange rates of glucose between blood and ISF and local blood flow around the sensor tip. Such effects can induce acute changes in the performance of the CGM system that lead to erroneous glucose readings or result in transient signal disruption but can last several hours.[55] Nontransient long-term changes

("sensor fouling") in glucose sensor performance have been described previously in detail.[56]

Impact of Algorithms and Filters Implemented in CGM Systems

It is important to understand that the CGM data used for MARD calculation are not the raw glucose signals measured by the glucose sensor, but those provided by the CGM system as output (stored data that can be downloaded for subsequent analysis). As the raw data, which are measured truly continuously, are superimposed with considerable amounts of noise and artifacts (with considerable differences between CGM systems depending on the sensor technology used), the data stream provided by the glucose sensor undergoes robust filtering and analysis by algorithms. Thereby the signal output is improved massively; however, it is an inherent feature of, for example, smoothing algorithms to introduce at least some physical time delay. Not much is known about the impact of algorithms and filtering on the MARD as the manufacturers of these devices regard these data and computational improvement opportunities as proprietary intellectual property, and the raw data, the sensor current, is rarely if ever available to outside parties or the end users.[3]

Estimation of MARD in Clinical Studies

The minimum performance criteria for BG systems are defined in the internationally accepted standard ISO 15197. The requirements for SMBG system accuracy are based on three considerations:

- the effectiveness of current technology for monitoring patients with diabetes mellitus
- recommendations of diabetes researchers as well as existing product standards and regulatory guidelines
- the state-of-the-art of BG monitoring technology

The advantage of having such a standard are:

- manufacturers know how good a newly developed BG system must be
- authorities know how good a BG system has to be in order to approve a new product
- clinicians and patients know how well they can trust the BG measurement results

As long as no internationally accepted standard for the performance of clinical evaluation studies of CGM systems is established and accepted by manufacturers, authorities and clinicians, each manufacturer can design and perform clinical studies according to its own discretion. Such a standard should also include recommendations for a structured and systematic evaluation of MARD; ideally in head-to-head studies. Without such an approach MARD numbers obtained by different studies have to be regarded with great care. The following aspects have to be taken into consideration:

Study Design

If the study design avoids large swings in glycemia, the MARD will be lower in comparison to the situation in which these show up. Ideally, two (or more) CGM systems would be studied in a head-to-head approach, as this neutralizes the potential impact of study cohort or study setting differences.

Patient Groups

Studies with patients with type 1 diabetes most probably will lead to higher MARD values than those with patients with type 2 diabetes as those with type 1 usually have larger swings in glycemia. Additionally, studies with well-controlled patients with a lower level of glycemic variability are expected to result in lower MARDs.

Reporting of Study Results

Besides providing an overall MARD (= mean of interindividual MARD values) as main study outcome, additional analyses should be performed providing MARD for time periods with different rates of glucose changes, different glucose ranges, night and day, duration of CGM usage. The MARD data obtained with all individual patients in a given clinical study should be presented in a sorted manner, listing the MARD by individual patient is one approach, or one can simply use the frequency distribution for the observed MARD within patients. This also provides information about the range of MARD results obtained with a given CGM system in different patients. The effect of compensation of measurement delays of the sensors should also be analyzed. This enables to construct a relation between the clinical study design and the accuracy of the computed MARD value. When two or more different CGM systems were studied in parallel, results of head-to-head assessments should also be presented. At least for some patients also intraindividual MARD data should be presented (if a given CGM system is used with multiple glucose sensors over time).

Such presentation of all MARD data available provide a better understanding of the CGM system performance and supports comparison of different CGM systems plus facilitates understanding the improvement seen with different generations of a given CGM system.

MARD Values of Different CGM Systems Obtained in Clinical Studies

For an optimal usability from a clinical point of view, CGM systems should provide glucose values that accurately reflect

BG values, that is, the glucose measurement should be performed with a high accuracy, good precision, and without outliers. One would assume that exactly this is evaluated in the multitude of clinical studies performed over time with different CGM systems (Table 3 and Figure 2). The range of MARD values seen with each of the different CGM systems listed can be attributed to the variability in study design, patient selection, comparison method, and so on used. However, this also implies that if the MARD and the way it is measured in clinical studies were a really reliable parameter, different studies would provide more or less the same MARD. In other words, by selecting certain study conditions (eg, more patients with type 2 diabetes), one can influence the MARD in the preferred direction, which is usually toward lower values. Without more standardized clinical trials, the MARD is not a reliable parameter from an analytical point of view.

Discussion and Conclusion

In view of the rapid improvements seen in the last 15 years with respect to the analytical performance of CGM systems, there is a clear need for a parameter that characterizes it adequately. Despite all its limitations, currently the overall MARD is the most often used parameter to characterize the analytical performance of CGM systems. This was appropriate when the MARD was established, as CGM systems were approved for "adjunctive use" only, that is, therapeutic decisions had to be based on additional (confirmation?) BG measurements. However, nowadays CGM systems are approved as "nonadjunctive"-systems. In this case use of the MARD as sole parameter to characterize the performance of CGM systems is not sufficient. For example, the overestimation of the performance using the MARD of a given CGM obtained if large swings in glycemia are avoided, is one clear example of such critical topics—it is hard to understand why the very same sensor is reported to have any MARD between 11% and 21%. As CGM systems are an essential part of each AID system, an in-depth evaluation of the performance of such CGM systems is critical.

From the regulative side, this has not yet happed, the one approved guideline for the evaluation of CGM systems (POCT-05 from CLSI; https://clsi.org/media/1502/poct05a_sample.pdf) that does exist is from 2008 and describes generic performance metrics and how studies should be designed and the data analyzed, but no acceptance criteria (like for BG systems) are given—a new version of this guideline is currently in preparation. The FDA has recently formulated criteria for a minimal performance of CGM systems that are used in combination with AID ("iCGM"; https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm602870.htm), but those are (so far) not binding for approval, but fulfilling them merely facilitates the approval process.

Against this background, attempts are going on to retain MARD, but to improve the information provided by it with calculation of additional parameters like MARDs computed for subsets of the entire dataset of the clinical trial (stratified by glucose range, time of sensor wear, etc), MARD reliability index (MRI[7]), or PARD.[5] Another option would be to define parameters that provide information about concentration errors (the MARD does not differentiate between accuracy and precision, see above) and also incorporate the additional information, such as trend arrows, that CGM provides.[57] Still, there is not yet a consensus which should be based on clinical relevance and therapeutic concepts into account as well, which is still missing.

Indeed, more insight is needed on several critical issues: what happens when glucose levels are shown x% too high and/or the trend arrow indicates an incorrect rate of change in glycemia in the near future? What are the clinical consequences of a therapeutic decision based on such incorrect information? Do these parameters—in particular MARD—reflect the daily experience of patients and health care professionals? If, say, CGM systems that are reported to have comparable MARD values, lead to a different clinical experience, the practical relevance of the metric is unclear.

The improvement of MARD values makes the picture even more complex. Indeed, the manufacturers proudly announce an improvement in MARD from one generation of their CGM system to the next, and the most recent generations of CGM systems are reported to have MARD values in the single digit range (<10%). How low will the MARD become in the future? 6%? will this analytical performance become as good as that of BG systems? But the key question: it is clear from the list of factors described above, that the minimum MARD that can be achieved will not be zero—if the comparison method is SMBG, MARD values of 5% would never be achievable. Also the other inherent measurement errors will remain with each CGM system and the comparison method. So the key question is: will a very low MARD be indicative at all of the accuracy of the device? Indeed, it is not even clear if a reduction in MARD from 13% to 10% reflects the same improvement in clinical outcome than a reduction from 10% to 7%.[58]

Summarizing, in view of the recent progress of the CGM systems, MARD needs to evolve as well, and in the meantime caution in its use is needed.

Abbreviations

AID, automated insulin delivery; BG, blood glucose; CGM, continuous glucose monitoring; ISF, interstitial fluid; MARD, mean absolute relative difference; MRI, MARD reliability index; PARD, precision absolute relative difference; SMBG, self-monitoring of blood glucose; TIR, time in range.

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: LH and GF support a number of companies in the development of novel diagnostic and therapeutic options for diabetes

treatment. MS, GSR, RH, and AK are employees of Roche Diabetes Care. FR and LDR have nothing to disclose.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Writing of the manuscript was supported by an unrestricted grant by Roche Diabetes Care.

ORCID iDs

Lutz Heinemann https://orcid.org/0000-0003-2493-1304
Günther Schmelzeisen-Redecker https://orcid.org/0000-0001-8915-7044
Guido Freckmann https://orcid.org/0000-0002-0406-9529

References

1. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. *Diabetes Care*. 2017;40:1631-1640.
2. Wilmoth DR. The relationships between common measures of glucose meter performance. *J Diabetes Sci Technol*. 2012;6:1087-1093.
3. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, Freckmann G, Heinemann L, Del Re L. Time delay of CGM sensors: relevance, causes, and countermeasures. *J Diabetes Sci Technol*. 2015;9:1006-1015.
4. Kirchsteiger H, Heinemann L, Freckmann G, et al. Performance comparison of CGM systems: MARD values are not always a reliable indicator of CGM system accuracy. *J Diabetes Sci Technol*. 2015;9:1030-1040.
5. Obermaier K, Schmelzeisen-Redeker G, Schoemaker M, et al. Performance evaluations of continuous glucose monitoring systems: precision absolute relative deviation is part of the assessment. *J Diabetes Sci Technol*. 2013;7:824-832.
6. Schoemaker M, Parkin C. CGM: how good is good enough? In: Kirchsteiger H, Jorgensen J, Renard E, del Re L, eds. *Prediction Methods for Blood Glucose Concentrations*. Cham, Switzerland: Springer; 2016:43-55.
7. 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.
8. Schrangl P, Reiterer F, Heinemann L, Freckmann G, Del Re L. Limits to the evaluation of the accuracy of continuous glucose monitoring systems by clinical trials. *Biosensors (Basel)*. 2018;8.
9. Rodbard D. Characterizing accuracy and precision of glucose sensors and meters. *J Diabetes Sci Technol*. 2014;8:980-985.
10. Noujaim SE, Horwitz D, Sharma M, Marhoul J. Accuracy requirements for a hypoglycemia detector: an analytical model to evaluate the effects of bias, precision, and rate of glucose change. *J Diabetes Sci Technol*. 2007;1:652-668.
11. 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;11:65-72.
12. 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;6:1103-1106.
13. 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;7:842-853.
14. Facchinetti A, Sparacino G, Guerra S, Luijf YM, DeVries JH, Mader JK, et al. Real-time improvement of continuous glucose monitoring accuracy: the smart sensor concept. *Diabetes Care*. 2013;36:793-800.
15. 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.
16. 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.
17. Facchinetti A, Del Favero S, Sparacino G, Cobelli C. Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices. *Med Biol Eng Comput*. 2015;53:1259-1269.
18. Luijf YM, Mader JK, Doll W, Pieber T, Farret A, Place J, et al. Accuracy and reliability of CGM systems: a head-to head comparison. *Diabetes Technol Ther*. 2013;15:721-726.
19. 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;7:833-841.
20. 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;7:1436-1445.
21. 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;8:699-708.
22. 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;16:759-767.
23. 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;17:343-349.
24. 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;9:801-807.
25. van Beers CA, DeVries JH, Kleijer SJ, et al. Continuous glucose monitoring for patients with type 1 diabetes and impaired awareness of hypoglycaemia (IN CONTROL): a randomised, open-label, crossover trial. *Lancet Diabetes Endocrinol*. 2016;4:893-902.
26. 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.
27. 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;10:876-884.
28. Bonora B, Maran A, Ciciliot S, et al. Head-to-head comparison between flash and continuous glucose monitoring

systems in outpatients with type 1 diabetes. *J Endocrinol Invest*. 2016;39:1391-1399.
29. 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;19:1051-1055.
30. Bailey TS, Chang A, Christiansen M. Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm. *J Diabetes Sci Technol*. 2015;9:209-214.
31. Freckmann G, Pleus S, Grady M, Setford S, Levy B. Measures of accuracy for continuous glucose monitoring and blood glucose monitoring devices [published online ahead of print November 19, 2018]. *J Diabetes Sci Technol*. doi:10.1177/1932296818812062.
32. 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;20:395-402.
33. Welsh JB, Kaufman FR, Lee SW. Accuracy of the Sof-sensor glucose sensor with the iPro calibration algorithm. *J Diabetes Sci Technol*. 2012;6:475-476.
34. 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;7:381-388.
35. Kovatchev B, Anderson S, Heinemann L, Clarke W. Comparison of the numerical and clinical accuracy of four continuous glucose monitors. *Diabetes Care*. 2008;31:1160-1164.
36. Diabetes Research in Children Network (DirecNet) Study Group; Buckingham B, Xing D, Weinzimer S, et al. Use of the DirecNet Applied Treatment Algorithm (DATA) for diabetes management with a real-time continuous glucose monitor (the FreeStyle Navigator). *Pediatr Diabetes*. 2008;9:142-147.
37. 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;14:225-231.
38. 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;15:758-761.
39. 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;7:842-853.
40. 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;16:667-678.
41. Bailey TS, Ahmann A, Brazg R, Christiansen M, Garg S, Watkins E, et al. Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. *Diabetes Technol Ther*. 2014;16:277-283.
42. Norgaard 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;9:215-220.
43. Christiansen MP, Garg SK, Brazg R, et al. Accuracy of a fourth-generation subcutaneous continuous glucose sensor. *Diabetes Technol Ther*. 2017;19:446-456.
44. 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;30:1125-1130.
45. Geoffrey M, Brazg R, Richard W. FreeStyle navigator continuous glucose monitoring system with TRUstart algorithm, a 1-hour warm-up time. *J Diabetes Sci Technol*. 2011;5:99-106.
46. Leelarathna L, Nodale M, Allen JM, et al. Evaluating the accuracy and large inaccuracy of two continuous glucose monitoring systems. *Diabetes Technol Ther*. 2013;15:143-149.
47. 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;17:801-807.
48. 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;17:763-765.
49. 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;40:63-68.
50. 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.
51. Nielsen JK, Freckmann G, Kapitza C, et al. Glucose monitoring by microdialysis: performance in a multicentre study. *Diabet Med*. 2009;26:714-721.
52. Siegmund T, Heinemann L, Kolassa R, Thomas A. Discrepancies between blood glucose and interstitial glucose-technological artifacts or physiology. *J Diabetes Sci Technol*. 2017;11:766-772.
53. Gehr B, Holder M, Kulzer B, et al. SPECTRUM: a training and treatment program for continuous glucose monitoring for all age groups [published online ahead of print August 20, 2016]. *J Diabetes Sci Technol*. doi:10.1177/1932296816661735.
54. 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;7:815-823.
55. Helton KL, Ratner BD, Wisniewski NA. Biomechanics of the sensor-tissue interface-effects of motion, pressure, and design on sensor performance and foreign body response-part II: examples and application. *J Diabetes Sci Technol*. 2011;5:647-656.
56. Anderson JM, Rodriguez A, Chang DT. Foreign body reaction to biomaterials. *Semin Immunol*. 2008;20:86-100.
57. Freckmann G, Link M, Westhoff A, Kamecke U, Pleus S, Haug C. Prediction quality of glucose trend indicators in two continuous tissue glucose monitoring systems. *Diabetes Technol Ther*. 2018;20:550-556.
58. Kovatchev BP, Patek SD, Ortiz EA, Breton MD. Assessing sensor accuracy for non-adjunct use of continuous glucose monitoring. *Diabetes Technol Ther*. 2015;17:177-186.