Journal of Diabetes Science and Technology
2023, Vol. 17(1) 133–140
© 2021 Diabetes Technology Society

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/19322968211039465
journals.sagepub.com/home/dst

Post-Market Surveillance Assessment of the Clinical Accuracy of a Blood Glucose Monitoring System with an Improved Algorithm for Enhanced Product Performance

Steven Setford, PhD[1], Zuifang Liu, PhD[1], David McColl, PhD[1], Stuart Phillips, MSc[1], Hilary Cameron, BSc[1], and Mike Grady, PhD[1]

Abstract

Background: On-going manufacturer-led post-market surveillance (PMS), assessing the clinical accuracy of blood glucose monitoring (BGM) systems, is critical to substantiate the performance of such products for people with diabetes.

Materials and Methods: Batches of Verio test-strip product were randomly and routinely selected over the period from launch of an improved-algorithm product to reporting date and sent to 3 clinic sites for clinician-led accuracy assessment. Accuracy is reported as per recently adopted FDA guidance for BGM systems, EN ISO 15197:2015 and MARD/MAD (Mean absolute relative difference/Mean absolute difference).

Results: Thirty-three individual test-strip batches were evaluated corresponding to 506 unique donors. Accuracy performance - FDA: 98.9% of values within ±15% of comparator; ISO: 99.0% within ±15 mg/dL or ±15% at <100 mg/dL (<5.55 mmol/L) or ≥100 mg/dL (≥5.55 mmol/L) glucose, respectively. Overall MARD was 4.19% with a MARD range of 3.54%-5.73% across all test strip batches.

Conclusions: This post-market surveillance program demonstrates the new BGM system consistently meets measures of clinical accuracy specified by regulators. This program supports a growing demand by regulators for real-world evidence demonstrating consistent in-market product efficacy as opposed to the current largely passive approach that relies on assessment of reports filed by device users.

Keywords

post-market surveillance (PMS), accuracy, MARD, blood glucose monitoring (BGM), regulations

Introduction

Blood glucose monitoring (BGM) systems remain critical tools for most people globally with diabetes who rely upon them to support diabetes management decisions where access to continuous glucose monitoring (CGM) systems is limited. Approval to market BGM products in the United States is governed by the Food and Drug Administration (FDA) who provide guidance regarding levels of performance for regulatory clearance.[1] From a user-evaluation perspective, BGMs must demonstrate 95% of all results to be within ±15% of the corresponding comparator result and 99% of results within ±20% of comparator across the claimed range. While not required by FDA guidance, MARD (Mean Absolute Relative Difference) provides another measure of accuracy, most widely quoted for CGM systems, but frequently used as an additional measure of BGM accuracy. To better meet end-user needs, BGM test-strip manufacturers strive to deliver new designs with improved clinical accuracy to market.

Regulatory submission of a BGM system is based on data from testing a relatively small number of batches of

[1]LifeScan Scotland Ltd., Inverness, UK

Corresponding Author:
Steven Setford, PhD, LifeScan Scotland Ltd, Beechwood Park North, Inverness, Highland IV2 3ED, UK.
Email: ssetford@lifescan.com

representative product, most likely manufactured within a short timeframe prior to that submission. However, the performance of test-strips should be continuously assessed over time (in-clinic) to ensure product quality and to demonstrate continued precision over the manufacturing process. Consequently, post-market surveillance (PMS) of BGM systems has been an ongoing area of concern,[2] with the Diabetes Technology Society (DTS) undertaking their own PMS exercise in 2016-2017 to assess the clinical performance of the 18 most sold BGMs in the U.S. market at that time.[3,4]

As commercially available products, the responsibility for continued assurance of clinical accuracy claims should lie with the legal manufacturer. LifeScan, as legal manufacturer of BGM systems routinely conducts such PMS.[5-8] Reported here is the accumulated surveillance data, since launch, of the U.S. version of the Verio Reflect BGM system, which incorporates a next-generation algorithm delivering improved clinical accuracy.

Methods

Test-strip

The Verio test-strip design incorporates co-facially opposed thin-film palladium and gold layers, separated by an insulating spacer, described in further detail elsewhere.[5] The palladium surface of the sample chamber (0.4 µL volume) is coated in reagent, that includes the enzyme FAD-GDH (flavin adenine dinucleotide-glucose dehydrogenase) that rapidly-solubilizes on blood application. A pulsed voltage-time waveform is applied during the 5 second test. The information-rich current-time transient response is algorithmically processed to give a hematocrit, temperature and interferent corrected response. The US Verio Reflect BGM system evaluated here employs an improved form of algorithm that further reduces the effect of electroactive interferences on system output.

Clinic Measurement

Clinical accuracy was assessed across 3 manufacturer-affiliated UK clinic sites: Highland Diabetes Institute, Inverness; Birmingham Heartlands Hospital, Birmingham; Royal Infirmary, Edinburgh. Each week, a minimum of one test-strip batch meeting the manufacturer's product release criteria was randomly selected and sampled for clinic evaluation purposes. In clinic, a target of 100 strips per batch were assessed on 100 individual subjects with capillary fingertip blood samples being drawn and tested as per manufacturer's instructions for use. Study inclusion criteria included a diagnosis of diabetes, a measured hematocrit of 20%-60% v/v and informed consent. No exclusions were made based on medications. Immediately following BGM measurement, 2 additional capillary samples were taken: A minimum 200 µL sample, collected into a 300 µL Microvette blood collection tube (Lithium:Heparin) for glucose determination by the comparator system. The time period between end of Microcuvette collection and start of centrifugation did not exceed 1 minute. An additional ~75 µL capillary blood was collected in a heparinized capillary tube for hematocrit measurement using a Hettich Hematocrit 210 capillary centrifuge.

All testing was facilitated by trained clinic staff. Following study completion, the validity of all datapoints were assessed against standard clinical acceptance criteria (hematocrit values within BGM claimed range, valid paired comparator value, adherence to protocol-time limits). The dataset is not entirely composed of unique donors since subjects may have visited the clinic site on more than one occasion over the evaluation period.

Comparator System

Microvette capillary blood samples were immediately centrifuged, with glucose content of the plasma fraction being determined, in duplicate, on 2 separate YSI 2300 STAT PLUS™ analyzers (Yellow Springs Instrument Co Inc, Yellow Springs, OH). The performance of the YSI 2300 (banks of 4 instruments) has previously been assessed in our laboratories using NIST standard reference material 965b (glucose in frozen human serum; levels 1, 2, 3, and 4). Pooled standard deviations, mean differences, mean errors and accuracy metrics were assessed and met the prescribed acceptance criteria.

Clinical Accuracy Calculations

The difference (bias) of a BGM value (test *i*) vs the corresponding comparator-derived value (YSI comparator *i*) was calculated as either absolute or relative difference values where:

$$\textit{Absolute Difference } i_{(BGM)} = \text{test } i - \text{YSI comparator } i$$

and

$$\textit{Relative Difference } i_{(BGM)} = \frac{\text{test } i - \text{YSI comparator } i}{\text{YSI comparator } i} \times 100\%$$

FDA accuracy analysis requires the percentage of relative difference values within ±15% or ±20% of comparator to be recorded, while EN ISO 15197:2015 accuracy analysis requires the percentage of values within either ±15 mg/dL (glucose <100 mg/dL) or ±15% (glucose ≥100 mg/dL) of comparator, respectively to be recorded.

MARD calculation: Firstly, the absolute (Abs) relative deviation (ARD) for a BGMS value (test *i*) and corresponding comparator value (YSI comparator *i*) was calculated:

$$ARD\ (\%) = \frac{Abs\left[\text{Difference}\left(\text{test } i, \text{YSI comparator } i\right)\right]}{\text{YSI comparator } i} \times 100\%$$

Mean absolute relative difference (MARD) from N paired test and comparator values was then calculated:

$$MARD(\%) = \frac{1}{N}\sum_{i=1}^{i=N} ARD$$

Absolute difference instead of ARD may be used at low glucose:

$$Absolute\ difference_{(glucose\ <100\ mg/dL)} = Difference\ (test\ i,\ YSI\ comparator\ i)$$

From which mean absolute deviation (MAD) may be determined. MAD is often quoted at low glucose as it provides a more readily interpretable quantitative measure of the difference between a test system and comparator.

Legacy-algorithm Performance Evaluation

The production-standard BGM meters used in clinic permitted post-measurement assessment of individual current-time transients gathered via the manufacturer's Medidata RAVE clinical registry (Medidata Solutions, NY). This allowed post-processing of the transient library using the both new algorithm and legacy algorithm code, facilitating direct comparisons between the next-generation and legacy-algorithm.

Results

Clinical Accuracy by FDA and EN ISO 15197:2015 Criteria

A total of 33 unique test-strip batches were evaluated equating to 3273 valid values from 506 unique donors. The distribution of clinic BG values, determined by comparator (Table 1) have a similar distribution to those seen in the manufacturer's previous clinic PMS datasets.[5-8]

Accuracy was assessed as per FDA and ISO 15197 accuracy definitions (Table 2), although it is to be appreciated that the methodology followed differed from current FDA and ISO 15197 accuracy assessment methodologies that require the user to perform the test. ISO 15197 has a further accuracy assessment termed "system accuracy," which permits a health care professional to perform the test but requires glucose values to be proportionally distributed within prescribed glucose ranges. However, this latter approach permits artificial manipulation of capillary blood samples to facilitate meeting the glucose distribution profile at extremes of glucose. At glucose levels <100 mg/dL, 99.4% (173/174) values were within ±15 mg/dL of comparator, while at glucose levels ≥ 100 mg/dL, 99.0% (3067/3099) values were within ±15%. A total of 99.6% of values (3261/3273) were within zone A of the Consensus Error Grid (CEG), with all other values within zone B. Figure 1 records meter relative difference values vs comparator values and indicates no systematic deviation in relative difference as a function of comparator glucose concentration with maximum recorded relative difference values of +28.3% and −26.3%.

Table 1. Comparator Method Glucose Value Distribution.

Glucose range (mg/dL)	Number of values	Percent (%)
≤50	10	0.3
50-≤80	40	1.2
80-≤120	320	9.8
120-≤200	1274	38.9
200-≤300	1094	33.4
300-≤400	422	12.9
>400	113	3.5
Total	3273	100.0

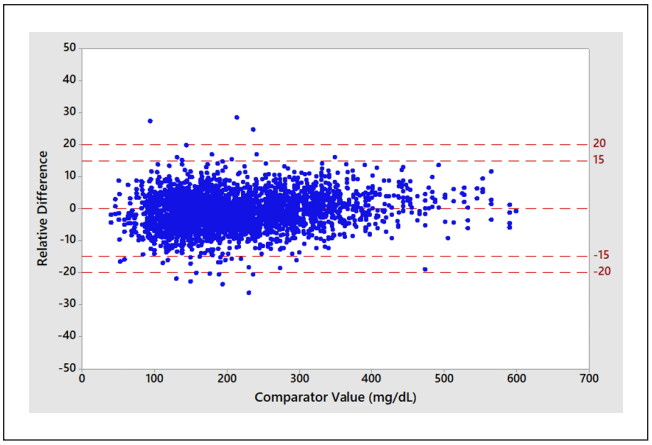

Figure 1. Individual meter results expressed as percent difference values to corresponding comparator value ($n = 3273$).

The Surveillance Error Grid (SEG) was developed to classify the risk posed by the magnitude of the inaccuracy of a glucose value (from a BGM or CGM) subsequently leading to an inappropriate treatment decision.[9,10] Eight levels of risk have been defined, ranging from "none" (risk level 0-0.5), through to extreme (>3.5). SEG analysis showed 99.3% (3249/3271) of the BGM values to be within the no risk zone, while 0.7% (22/3271) of values scored in the slight risk (lower) zone (Figure 2). No values were recorded in any of the other risk zones.

Clinical Accuracy by MARD and MAD

The MARD for the full clinic dataset of 33 strip batches was 4.19% (Table 3). Applying an ISO 15197 equivalent glucose

Table 2. Test-strip Performance vs FDA and ISO 15197: 2015 Clinical Accuracy Criteria.

Criterion	Sample size (*n*)	Within accuracy criterion (*n*)	Within accuracy criterion (%)	Lower confidence interval (%)	Pass/fail
FDA (±15%)[1]	3273	3238	98.9	98.6	Pass
FDA (±20%)[2]	3273	3262	99.7	99.4	Pass
ISO (±15 mg/dL or 15%)[3]	3273	3240	99.0	98.7	Pass
Zone A + B of consensus error grid[4]	3273	3273	100.0	na	Pass

[1]95% of results to be within ±15% of corresponding comparator result.
[2]99% of results to be within ±20% of comparator result.
[3]95% of results to be within ±15 mg/dL (0.83 mmol/L) of corresponding comparator result at glucose values <100 mg/dL (5.55 mmol/L) or 15% of comparator result at glucose ≥100 mg/dL (5.55 mmol/L).
[4]At least 99% of values, to be within zones A + B of the Consensus Error Grid (CEG).

cut-off criterion to the dataset yielded a MARD of 4.17% (glucose ≥100 mg/dL) and a MAD of 3.88 mg/dL (glucose <100 mg/dL). MAD was also calculated for the 33 values within the hypoglycaemic range (<70 mg/dL) where a mean absolute deviation of 2.58 mg/dL was recorded. To quantify batch-to-batch differences, MARD was calculated by individual batch, ranging from 3.54%-5.73% across the 33 batches.

Next-Generation Algorithm vs Legacy-algorithm

The next-generation algorithm gave 98.9% of values within ±15% of comparator (lower confidence interval [LCI] 98.6%), a higher figure when compared to the legacy-algorithm (97.6% [97.1%]). Corresponding figures within ±20% of comparator for next-generation and legacy algorithms were: 99.7% [99.4%] and 99.5% [99.2%], respectively. Against the ISO 15197-derived criterion of ±15 mg/dL/±15%, 99.0% [98.7%], and 97.6% [97.2%] of values met the requirement for next-generation and legacy, respectively. Both algorithms gave 100% of values within zones A and B of the CEG, with 12 and 19 values within zone B for the next-generation and legacy algorithm, respectively.

An improvement in MARD performance was similarly observed by application of the improved algorithm, with reductions in overall MARD from 5.17% to 4.19% and MARD (≥ 100 mg/dL glucose) from 5.10% to 4.17%. Similar reductions in MAD were also recorded, falling from 5.28 mg/dL to 3.88 mg/dL (<100 mg/dL glucose) and 3.51 mg/dL to 2.58 mg/dL (<70 mg/dL). The range of MARDs by batch for the next-generation algorithm was 3.54%-5.73%, an improvement over legacy-algorithm performance of 4.37%-6.06%.

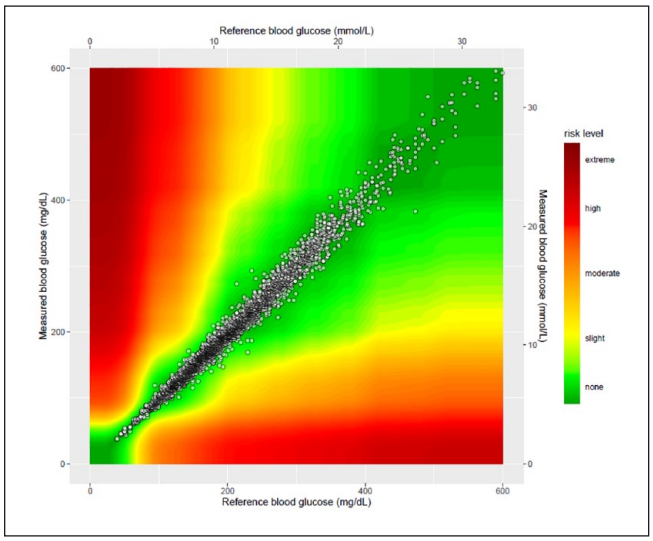

Figure 2. Surveillance Error grid (SEG) plot of individual meter glucose value vs corresponding comparator value (*n* = 3271). Note: Two values from the n = 3273 dataset could not be included in SEG analysis since while the comparator values were below the 600 mg/dL cut-off, the corresponding meter values were >600 mg/dL.

Discussion

Accuracy

The BGM subjected to this on-going post-market surveillance (PMS) meets clinical accuracy requirements, as set out by current FDA guidance. This HCP-led approach represented the most pragmatic means for the manufacturer to obtain clinical performance data from finger-tip capillary bloods on an on-going basis for multiple strip-batches and BGM platforms. For the record, the performance of this BGM system in terms of lay-user evaluation, clinical validation and extreme glucose evaluation has been reported elsewhere.[11] This dataset, comprising 3273 paired BGM-comparator values, represents an approximately 10-fold increase in the paired data-points required for an FDA lay-user clinical verification study. As this PMS activity continues, the growing dataset will be used to continually assess product performance.

MARD

The FDA approach of reporting clinical accuracy in terms of percentage of results within a given relative value of the corresponding comparator value is binary: values are either

Table 3. Clinical PMS Dataset Expressed by Mean Absolute Relative Deviation (MARD) and Mean Absolute Deviation (MAD) Across Different Glucose Ranges.

Metric	Glucose range (mg/dL)	MARD or MAD value	Number of paired data values
MARD	39-599	4.19%	3273
MARD	≥100	4.17%	3099
MAD	<100	3.88 mg/dL	174
MAD	<70	2.58 mg/dL	33

within or not within the acceptance threshold. MARD offers an alternative measure of accuracy, serving to quantify the extent of the mean deviation of a set of test system results from the comparator method. That is not to say that MARD alone should be an indicator of accuracy, since it is non-directional in terms of indicating a positive or negative bias to comparator. Furthermore, MARD has been shown to be influenced by multiple factors[12,13] including glucose concentration and time of day of measurement.[14]

The overall MARD of 4.19% (range 3.54-5.73) recorded in this study compared well to leading BGM products derived from independent evaluations.[15] While MARD does not differentiate between systematic errors (bias) and random errors (precision), a product with a consistently low batch MARD is indicative of a BGM with a relatively narrow systematic error. Bedini et al.[16] have modelled the effect of hypothetical bolus insulin dosing errors based on BGM MARD accuracy, concluding that systems with lowest MARD reduce the likelihood of clinically significant events. Furthermore, a product with lower MARD is more beneficial since it would manifest itself, in the hands of the end-user, as a less perceptible change in reading as the user starts a new vial of strips from a new strip batch.

The SEG risk levels are such that, at low glucose (<100 mg/dL), a small absolute (mg/dL) inaccuracy in a BGM value can result in a significant shift in risk level. While points plotted on the SEG represent individual BGM-comparator pairings, the overall MAD values of 3.88 mg/dL ($n=174$) and 2.58 mg/dL ($n=33$) for glucose values <100 and <70 mg/dL respectively indicate how BGM performance supports diabetes management decisions within this sensitive low euglycaemic/hypoglycaemic range.

Several efforts have been made to relate MARD values to the likelihood of meeting ISO performance. Pardo and Simmons[17] explored the relationship between MARD and ISO, concluding that batches with MARDs of 4.25% were "most likely" to meet the ISO standard, with values of 3.25% and 5.25% representing the "most stringent" and "most liberal" conditions, respectively). However, this analysis was based on a modelling approach. An empirical evaluation by Freckmann et al.[18] summarizing the results of 77 separate system accuracy studies (169 BGMs, 809 test-strip lots) performed in their laboratory found that all strip batches with a MARD of ≤6.1% met EN ISO 15197:2015, while batches with MARD of ≤4.1% yielded performance equivalent to ±10 mg/dL/±10% ISO accuracy.

Next Generation Algorithm

Most BGM systems use redox enzymes such as glucose oxidase or glucose dehydrogenase as specific biorecognition agents to oxidize glucose in the sample with transfer of electrons facilitated by reduction of a mediator. For electrochemical BGMs, the reduced mediator is re-oxidized at a suitably positively charged working electrode. However, certain compounds in blood, whether exogenous or endogenous, may also undergo oxidation at the electrode, adding a non-glucose specific contribution to the raw signal. BGM or CGM devices (whether using blood or interstitial fluid), will contain endogenous interferences, such as uric acid, and may contain exogenous interferences, notably ascorbic acid or acetaminophen, that are electroactive and so may also affect the reported glucose value.

The next generation algorithm, first implemented in the BGM system reported here, exploits the pulsed nature of the potential-time waveform applied to the test-strip to further desensitize the system to electroactive interferences. The potentials of the 2-electrode design are switched relative to each other during the 5 s assay, with the raw-signal contribution of any electroactive interfering species present being determined by assessing the relative magnitudes of response between the reagent coated and non-reagent coated electrodes. The outcome of this algorithmic enhancement is evidenced by the FDA clinical accuracy and MARD improvements recorded (98.9% vs 97.6% within ±15% of comparator; 5.17 vs 4.19% vs 5.17% MARD).

Sponsor Bias in BGM Studies

An in-depth comparison of the clinical accuracy obtained in this study against published accuracy for other BGMs is not included. Such an analysis would present multiple datasets from multiple investigators with often conflicting data regarding product performance. In some cases, the assessment method may be responsible for differences observed, such as choice of comparator employed by the manufacturer for product release and that used by a researcher for product assessment.[19] Furthermore the concept of sponsor-bias cannot be ignored.[20] This phenomenon relates to studies in which the sponsor's product performs favourably against carefully selected competitor systems, while independently funded studies, undertaken on the same products, head-to-head, can result in different outcomes. This may be evidenced by comparing results from BGMs investigated within

the DTS surveillance program vs sponsor-supported evaluations of the same products. Crucially, while it is recognized that independent investigators will undertake a sponsor-funded evaluation with complete integrity, the study sponsor themselves may often reserve the right to choose whether to publish the data (this is not the case for investigator-initiated studies or in situations where the sponsor provides unrestricted funds, giving the researcher freedom to publish their findings). The opportunity to withhold unfavourable data is eliminated if a manufacturer adopts and adheres to a routine PMS program.

Glucose test strips are manufactured on a batch basis, with the goal of producing batches with consistent clinical performance. However, the mean bias between batches varies due to differences in raw materials, strip stability and manufacturing processes and conditions. Batches that more closely agree with the comparator (or reference) performance, that is, have a mean bias closer to 0% will generally exhibit high accuracy. Therefore, a batch with a mean bias of ~0% will more likely out-perform a batch with a bias closer to the outer limits of the release conditions. This generalized pattern may be seen in Freckmann et al.'s analysis, where only a small proportion of strip batches exhibiting centred biases failed to meet the ISO standard (however, it should also be noted that a significantly non-centred batch can also met the standard should the batch exhibit tight precision).[18] A properly constructed and ongoing PMS program, based on an assumption that production batches are truly sampled randomly and that data from all tested batches is reported, removes the opportunity for a sponsor to pre-select a batch for assessment based on known performance, and subsequently decide whether to publish the findings based on a retrospective evaluation of the results.

Benefit of a BGM Post-market Surveillance (PMS) Program

The demand remains for next generation BGMs that deliver improved clinical accuracy. These systems are needed across diverse global markets where CGM adoption remains low due to cost, availability and support. Many CGM systems still require frequent calibration, in turn requiring reliable, high accuracy BGMs to maximize clinical performance and reduce risk of detrimental insulin dosing decisions.[21,22] CGM-users are also encouraged to compare values to BGMs in situations where CGM values do not match symptoms.[23] Klonoff[24] has recently commented on a case where a multicentre drug study required protocol revision due to safety concerns regarding the accuracy of an FDA-cleared BGM and has called for PMS of devices for the safety of subjects and accurate determination of the effectiveness of diabetes drugs and devices. Others also argue that current standards of BGM and indeed CGM performance are inadequate, proposing monitoring the performance of devices in the field.[25]

Regulators and leading commentators are actively calling for PMS activities.[24-27] The FDA advocates moving from the current largely passive approach of device users reporting issues to a more active real-world evidence-based approach.[28] Similarly, recent changes to the Medical Device Regulation (MDR) and In Vitro Diagnostic Regulation (IVDR) in Europe have led to a substantial change in legislation requiring more thorough oversight by manufacturers of medical devices once in market.[29]

A properly constructed BGM (and future CGM) PMS programme will ensure that production batch selection is random, frequent, on-going and made available for public scrutiny, eliminating the opportunity to supress undesirable performance data. In this study, the manufacturer/sponsor followed a defined standard operating procedure to ensure appropriate random selection of batches, sampling frequency and study execution and was undertaken at 3 independent clinics, per the recommendations of the diabetes technology society (DTS).

Conclusions

While regulators enforce standards that ensure BGMs meet fundamental requirements, there is no absolute requirement for manufacturers to demonstrate that their products consistently maintain those same standards of performance. Traditionally, independent researchers, or interested industry bodies such as the Diabetes Technology Society, do test BGMs against their claims and publish the findings. However, such activities are costly, often with logistical issues in obtaining representative quantities of test materials and are generally limited to a point in time evaluation with limited numbers of test-strip batches. We contend that the responsibility to ensure that BGMs continue to meet their claims lies not with researchers or societies, but with manufacturers. Reported here is just such a longitudinal study based on a novel high-accuracy BGM system deploying an improved algorithm design, demonstrating both consistent batch performance against recent FDA guidance for BGMs and a clear improvement in clinical accuracy based on an improved algorithm design.

Abbreviations

BGM, Blood Glucose Monitoring; CEG, Consensus Error Grid; CGM, Continuous Glucose Monitoring; DTS, Diabetes Technology Society; FDA, Food and Drug Administration; ISO, International Standards Organisation; IVDR, In Vitro Devices Regulation; LCI, Lower Confidence Interval; MAD, Mean absolute Difference; MARD, Mean Absolute Relative Difference; MDR, Medical Device Regulation; PMS, Post-Market Surveillance; SEG, Surveillance Error Grid; YSI, Yellow Springs Instruments.

Acknowledgments

Grateful thanks to all staff at LifeScan responsible for the development of the improved algorithm, the LifeScan clinical team and clinical staff at each of the test sites for undertaking these PMS activities.

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: Steven Setford, Zuifang Liu, David McColl, Stuart Phillips, Hilary Cameron, and Mike Grady are full time employees of LifeScan Scotland Ltd.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by LifeScan Scotland Ltd.

ORCID iDs

Steven Setford https://orcid.org/0000-0003-1983-0461
Stuart Phillips https://orcid.org/0000-0003-2399-9016
Mike Grady https://orcid.org/0000-0002-8006-1844

References

1. Food and Administration Drug. *Self-Monitoring Blood Glucose Test Systems for Over-the-Counter Use. Guidance for Industry and Food and Drug Administration Staff*. United States Food and Drug Administration (FDA); 2020.
2. Klonoff DC, Prahalad P. Performance of cleared blood glucose monitors. *J Diabetes Sci Technol*. 2015;9(4):895-910.
3. Klonoff DC, Parkes JL, Kovatchev BP, et al. Investigation of the accuracy of 18 marketed blood glucose monitors. *Diabetes Care*. 2018;41(8):1681-1688.
4. Klonoff DC, Lias C, Beck S, et al. Development of the Diabetes Technology Society blood glucose monitor system surveillance protocol. *J Diabetes Sci Technol*. 2016;10(3):697-707.
5. Setford S, Grady M, Phillips S, et al. seven-year surveillance of the clinical performance of a blood glucose test strip product. *J Diabetes Sci Technol*. 2017;11(6):1155-1162.
6. Setford S, Grady M, Mackintosh S, Donald R, Levy B. Seven-year clinical surveillance program demonstrates consistent MARD accuracy performance of a blood glucose test strip. *J Diabetes Sci Technol*. 2018;12(5):1016-1023.
7. Setford S, Phillips S, Grady M. Evidence from a long-term, systematic post-market surveillance program: clinical performance of a hematocrit-insensitive blood glucose test strip. *J Diabetes Sci Technol*. 2021;15(1):67-75.
8. Grady M, Cameron H, Phillips S, Smith G, Setford S. Analysis of a unique postmarket surveillance dataset that a glucose test-strip demonstrates no evidence of interference and robust clinical accuracy irrespective of the prescription medication status of a large cohort of patients with Diabetes. *J Diabetes Sci Technol*. 2021;15(1):82-90.
9. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014;8(4):658-672.
10. Kovatchev BP, Wakeman CA, Breton MD, et al. Computing the surveillance error grid analysis: procedure and examples. *J Diabetes Sci Technol*. 2014;8(4):673-684.
11. Katz LB, Smith A, Cameron H, Setford S, Grady M. 10/10 accuracy analysis of a self-monitoring blood glucose system. *J Diabetes Sci Technol*. 2021;15(3):717-718.
12. Heinemann L, Schoemaker M, Schmelzeisen-Redecker G, et al. Benefits and limitations of MARD as a performance parameter for continuous glucose monitoring in the interstitial space. *J Diabetes Sci Technol*. 2020;14(1):135-150.
13. Freckmann G, Pleus S, Grady M, Setford S, Levy B. Measures of accuracy for continuous glucose monitoring and blood glucose monitoring devices. *J Diabetes Sci Technol*. 2019;13(3):575-583.
14. Pleus S, Stuhr A, Link M, Haug C, Freckmann G. Variation of mean absolute relative differences of continuous glucose monitoring systems throughout the day. *J Diabetes Sci Technol*. Published online February 20, 2021. doi:10.1177/1932296821992373
15. Ekhlaspour L, Mondesir D, Lautsch N, et al. Comparative accuracy of 17 point-of-care glucose meters. *J Diabetes Sci Technol*. 2017;11(3):558-566.
16. Bedini JL, Wallace JF, Pardo S, Petruschke T. Performance evaluation of three blood glucose monitoring systems using ISO 15197: 2013 accuracy criteria, consensus and surveillance error grid analyses, and insulin dosing error modeling in a hospital setting. *J Diabetes Sci Technol*. 2016;10(1):85-92.
17. Pardo S, Simmons DA. The quantitative relationship between ISO 15197 accuracy criteria and mean absolute relative difference (MARD) in the evaluation of analytical performance of self-monitoring of blood glucose (SMBG) systems. *J Diabetes Sci Technol*. 2016;10(5):1182-1187.
18. Freckmann G, Mende J, Pleus S, et al. Mean absolute relative difference of blood glucose monitoring systems and relationship to ISO 15197. *J Diabetes Sci Technol*. Published online March 24, 2021. doi:10.1177/19322968211001402
19. Pleus S, Baumstark A, Jendrike N, Mende J, Schauer S, Haug C, et al. Impact of two different comparison methods on apparent system accuracy of 18 CE-marked current-generation self-monitoring blood glucose systems. Poster #104, 13th International Conference on Advanced Technologies and Treatments for Diabetes; February 19-22, 2020; Madrid. Mary Ann Liebert, Inc.
20. Krouwer JS. Biases in clinical trials performed for regulatory approval. *Accreditation Qual Assur*. 2015;20:437-439.
21. Breton MD, Kovatchev BP. Impact of blood glucose self-monitoring errors on glucose variability, risk for hypoglycemia, and average glucose control in type 1 diabetes: AnIn SilicoStudy. *J Diabetes Sci Technol*. 2010;4(3):562-570.
22. Heinemann L. Continuous glucose monitoring (CGM) or blood glucose monitoring (BGM): interactions and implications. *J Diabetes Sci Technol*. 2018;12(4):873-879.
23. Funtanilla VD, Candidate P, Caliendo T, Hilas O. Continuous glucose monitoring: a review of available systems. *P T*. 2019;44(9):550-553.
24. Klonoff DC. Postmarket surveillance of blood glucose monitor systems Is needed for safety of subjects and accurate determination

of effectiveness in clinical trials of Diabetes drugs and devices. *J Diabetes Sci Technol*. 2019;13(3):419-423.
25. Krouwer JS, Cembrowski GS. A review of standards and statistics used to describe blood glucose monitor performance. *J Diabetes Sci Technol*. 2010;4(1):75-83.
26. US Food and DrugD Administration. Strengthening our national system for medical device postmarket surveillance center for devices and radiological health. 2012. Accessed April 14, 2021. http://wayback.archive-it.org/7993/20170112001851/http:/www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM301924.pdf
27. Klonoff DC. The expanding role of real-world evidence trials in health care decision making. *J Diabetes Sci Technol*. 2020;14(1):174-179.
28. US Food and DrugD Administration. Leveraging real world evidence in regulatory submissions of medical devices. In: Shuren J, Canos D, eds. Accessed March 16, 2021. https://www.fda.gov/news-events/fda-voices/leveraging-real-world-evidence-regulatory-submissions-medical-devices
29. Pane J, Francisca RDC, Verhamme KMC, et al. EU postmarket surveillance plans for medical devices. *Pharmacoepidemiol Drug Saf*. 2019;28:1155-1165.