Investigation of the Accuracy of 18 Marketed Blood Glucose Monitors

David C. Klonoff,[1] Joan Lee Parkes,[2] Boris P. Kovatchev,[3] David Kerr,[4] Wendy C. Bevier,[4] Ronald L. Brazg,[5] Mark Christiansen,[6] Timothy S. Bailey,[7] James H. Nichols,[8] and Michael A. Kohn[9]

Diabetes Care 2018;41:1681–1688 | https://doi.org/10.2337/dc17-1960

OBJECTIVE

Cleared blood glucose monitors (BGMs) for personal use may not always deliver levels of accuracy currently specified by international and U.S. regulatory bodies. This study's objective was to assess the accuracy of 18 such systems cleared by the U.S. Food and Drug Administration representing approximately 90% of commercially available systems used from 2013 to 2015.

RESEARCH DESIGN AND METHODS

A total of 1,035 subjects were recruited to have a capillary blood glucose (BG) level measured on six different systems and a reference capillary sample prepared for plasma testing at a reference laboratory. Products were obtained from consumer outlets and tested in three triple-blinded studies. Each of the three participating clinical sites tested a different set of six systems for each of the three studies in a round-robin. In each study, on average, a BGM was tested on 115 subjects. A compliant BG result was defined as within 15% of a reference plasma value (for BG ≥100 mg/dL [5.55 mmol/L]) or within 15 mg/dL (0.83 mmol/L) (for BG <100 mg/dL [5.55 mmol/L]). The proportion of compliant readings in each study was compared against a predetermined accuracy standard similar to, but more lenient than, current regulatory standards. Other metrics of accuracy included the overall compliance proportion; the proportion of extreme outlier readings differing from the reference value by >20%; modified Bland-Altman analysis including average bias, coefficient of variation, and 95% limits of agreement; and proportion of readings with no clinical risk as determined by the Surveillance Error Grid.

RESULTS

The different accuracy metrics produced almost identical BGM rankings. Six of the 18 systems met the predetermined accuracy standard in all three studies, 5 systems met it in two studies, and 3 met it in one study. Four BGMs did not meet the accuracy standard in any of the three studies.

CONCLUSIONS

Cleared BGMs do not always meet the level of analytical accuracy currently required for regulatory clearance. This information could assist patients, professionals, and payers in choosing products and regulators in evaluating postclearance performance.

[1]Mills-Peninsula Medical Center, Diabetes Research Institute, San Mateo, CA
[2]Joan Lee Parkes Consulting, Inc., Bristol, IN
[3]Center for Diabetes Technology, University of Virginia, Charlottesville, VA
[4]William Sansum Diabetes Center, Santa Barbara, CA
[5]Rainier Clinical Research Center, Inc., Renton, WA
[6]Diablo Clinical Research, Walnut Creek, CA
[7]AMCR Institute, Inc., Escondido, CA
[8]School of Medicine, Vanderbilt University, Nashville, TN
[9]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA

Corresponding author: David C. Klonoff, dklonoff@diabetestechnology.org.

Received 19 September 2017 and accepted 3 May 2018.

Clinical trial reg. no. NCT02789319, clinicaltrials.gov.

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc17-1960/-/DC1.

This article is featured in a podcast available at http://www.diabetesjournals.org/content/diabetes-core-update-podcasts.

Self-testing of blood glucose (BG) using a personal blood glucose monitor (BGM) is a cornerstone of diabetes treatment (1). BGMs are used for *1*) measuring BG to determine therapeutic decisions, *2*) calibrating continuous glucose monitoring systems, and *3*) detection or confirmation of hypoglycemia. To be both safe and of clinical value, BGM systems should measure BG levels accurately (2).

International Organization for Standardization (ISO) 15197:2013 is an international standard for defining the accuracy of BGMs (3). However, it is not used by the U.S. Food and Drug Administration (FDA) as part of the clearance process for these devices. In 2016, the FDA developed a standard for BGMs for over-the-counter use (4) that was similar to ISO 15197:2013. In defining an acceptable level of accuracy, these two standards both require 95% of data pairs (defined as a BGM measurement and a reference measurement) to be within 15% for BG values >100 mg/dL (5.55 mmol/L). However, for BG values <100 mg/dL (5.55 mmol/L), ISO 15197:2013 requires data pairs to be within 15 mg/dL (0.83 mmol/L), whereas the FDA 2016 over-the-counter standard requires data pairs to be within 15%. These two standards also differ in the number of data pairs required for testing and the acceptable number of extreme outlier data pairs.

In recent years, personal BGMs have been reported to perform below international standards or FDA standards (5–7). Furthermore, adverse clinical and economic outcomes due to analytical inaccuracy of BGMs have been reported through empirical and modeling studies (5). Inaccurate BGMs could potentially put users at significant personal risk. The Diabetes Technology Society (DTS)-BGM Surveillance Program for marketed BGMs was developed to provide an independent assessment of the analytical performance of BGMs after clearance by the FDA and also to provide information that can assist the diabetes community, health care professionals, and payers to make informed decisions when selecting a BGM (8). The consensus protocol that was used in this study was developed by a panel of experts in clinical chemistry, clinical diabetes, and regulatory science with representation from the FDA, Centers for Disease Control and Prevention, National Institutes of Health, U.S. Army, professional organizations, academia, and industry.

RESEARCH DESIGN AND METHODS

This study was conducted during 2016–2017. The study was approved by the Schulman Institutional Review Board, and all subjects gave informed consent prior to participating in the study. The 18 selected BGMs (Table 1) represented the best-selling BGMs in the U.S., which comprised approximately 90% of products obtained from consumer outlets between 2013 and 2015 as measured by IMS Rx Xponent data (9), Office of Inspector General's Medicare mail-order survey (10), and private label products (11). These BGMs were tested at three clinical sites (Rainier Clinical Research Center, Inc.; Diablo Clinical Research, Inc.; and AMCR [Advanced Metabolic Care + Research] Institute, Inc.). Every subject had a capillary BG level measured on six different BGMs and a simultaneous reference capillary sample prepared for comparator plasma testing at a reference laboratory. Each of three separate studies tested all 18 BGMs. Each clinical site assessed all 18 systems by testing a different set of six BGMs for each of the three studies in a round-robin. The reference laboratory was William Sansum Diabetes Center.

Subjects

Enrolled subjects were aged 18 years and older and had type 1 diabetes, type 2 diabetes, prediabetes, or no diabetes. Exclusion criteria included: *1*) hemophilia or any other bleeding disorder; *2*) pregnancy; and *3*) a condition, which in the opinion of the investigator or designee, would put the person or study at risk. Subjects completed initial screening to assess eligibility. Limited demographic and medical information about the subjects was collected including age, sex, race, ethnicity, presence/type of diabetes, fasting or not fasting state, and medications, although this information was not used for inclusion or exclusion.

Table 1—List of 18 BGMs (monitor name, strip name, and manufacturer) with measuring range for glucose and hematocrit range limit

BGM name	Strip name	Manufacturer	Glucose range (mg/dL)	HCT range (%)
Accu-Chek Aviva Plus	Accu-Chek Aviva Plus	Roche	20–600	10–65
Accu-Chek SmartView	Accu-Chek SmartView	Roche	20–600	10–65
Advocate Redi-Code +	Advocate	Diabetic Supply of Suncoast	20–600	20–60
Contour Classic	Contour	Bayer	10–600	0–70
Contour Next	Contour Next	Bayer	20–600	15–65
CVS Advanced	CVS Advanced	AgaMatrix	20–600	20–60
Embrace	Embrace No-Code	Omnis Health	20–600	30–55
FreeStyle Lite	FreeStyle Lite	Abbott Diabetes Care	20–500	15–65
Gmate SMART	Gmate	Philosys, Inc.	20–600	20–60
OneTouch Ultra 2	OneTouch Ultra	LifeScan	20–600	30–55
OneTouch Verio	OneTouch Verio	LifeScan	20–600	20–60
Prodigy AutoCode	Prodigy No Coding	Prodigy	20–600	30–55
Solus V2	Solus	BioSense Medical	20–600	32–56
TRUEresult	TRUEresult	HDI/Nipro	20–600	20–55
TRUEtrack	TRUEtrack	HDI/Nipro	20–600	30–55
Walmart ReliOn Prime	ReliOn Prime	ARKRAY	20–600	33–52
Walmart ReliOn Confirm (Micro)	ReliOn Confirm/Micro	ARKRAY	20–600	30–54
Walmart ReliOn Ultima	ReliOn Ultima	Abbott Diabetes Care	20–500	30–60

A sample size of approximately 100 subjects in each study of each BGM was recommended by the surveillance protocol as large enough to adequately assess accuracy.

Sources of BGMs and Strips

The BGMs and test strips were obtained from various parts of the U.S. both from large retail pharmacies and from online retailers to mimic the experience of people with diabetes. There was no requirement regarding how many test strip lots were to be used per study or in the overall project.

Testing Procedure

Recruitment was designed to obtain a wide range of BG values. For enhancement of the percentage of low BG values, approximately one-third of recruited subjects did not have diabetes and some were asked to come to the clinic fasting. Conversely, for obtainment of values in the high glucose range, some subjects with diabetes were tested 60–120 min after a meal.

In each study at each site, every subject had finger-stick capillary blood obtained and measured on six BGMs. The sequence of testing six BGMs was randomized for each subject by a predetermined schedule that assigned all BGMs to be equally frequently in all six positions at each site. The BGM display readings, including any error messages, were recorded and photographed. Reference plasma samples were obtained via deep finger puncture and collected into a tube containing lithium heparin. This blood was centrifuged within 5 min of collection and the plasma was transferred to a tube without additive, frozen, and later shipped to the reference laboratory for analysis.

The deep stick to obtain a blood sample for reference testing was performed in the middle of the sequence of six finger-sticks for BGM testing. All six finger-sticks were done within 12 min of the deep stick to minimize in vivo changes in BG concentration during the testing period. All tests were performed by trained staff who were health care professionals. Subjects did not perform any self-testing.

This study was not designed or powered to look at glycemic bin subsets, such as in the range of $<$70 mg/dL (3.9 mmol/L) or $>$300 mg/dL (16.6 mmol/L). We also tested glycolized specimens on a subset of the 18 BGMs as part of this protocol and will report those results separately.

Capillary plasma was shipped to the reference laboratory for testing on five comparator instruments (YSI Life Sciences 2300 Stat Plus Glucose Lactate Analyzer). The comparator instruments performed autocalibration every 20 min and underwent regular stringent quality control testing, including testing with four glucose levels of National Institute of Standards and Technology (NIST) Standard Reference Material 965b Standards (Glucose in Frozen Serum) to assure accuracy traceable to mass spectrometry measurement.

Blinding

This study was triple blinded. Those reading the BGMs could not know the plasma reference measurements, since they were performed later and at a different location; the reference laboratory did not have the BGM readings, and during the collation and analysis of the results, the BGMs were designated by a code number. All analyses were performed while blinded to the identity of the meters. The results and conclusions were posted prior to unbinding. The sponsor was not aware of the data and was not involved in the writing of the manuscript.

Analyses

The protocol specified that to be compliant a BGM's reported value must be within 15% of a reference plasma value $\geq$100 mg/dL (5.55 mmol/L) and within 15 mg/dL (0.83 mmol/L) of a reference value $<$100 mg/dL (5.55 mmol/L). This definition of data point compliance is the same as that used by ISO 15197:2013, which requires that 95% of a study's data pairs be compliant for a BGM to pass (3). The FDA's over-the-counter 2016 standard had not been released when the surveillance protocol was developed. This FDA guidance requires 95% of all BG results to be within 15% of the comparator results across the entire claimed measuring range of the device and that 99% of all BGM results be within 20% of the comparator results across the entire claimed measuring range of the device (4), which is a more stringent requirement than that specified by ISO 15197:2013 (3).

We applied a predetermined accuracy standard to each BGM in each of the three studies. A BGM was considered to have met the standard unless the proportion compliant was in the "clear rejection zone" (7). The clear rejection zone is defined by taking as the null hypothesis that the meter meets the ISO 15197:2013 level of performance, which is 95% compliant readings. For rejection of this hypothesis, the measured number of compliant readings must be low enough such that chance variation would account for this poor outcome $<$5% of the time. With 100 readings, the number compliant to reject the null hypothesis must be $<$91 (91% compliant) or the BGM meets the accuracy standard. With 125 readings, the number compliant to reject the null hypothesis must be $<$115 (92% compliant) or the BGM meets the accuracy standard. This means that a device with a compliant proportion as low as 91 of 100 (91%) or 115 of 125 (92%) would meet the accuracy standard, making it more lenient than the ISO standard of 95% compliance.

Four other metrics of accuracy were also assessed. These metrics included *1*) overall compliance in the three studies (total compliant readings/total readings); *2*) number and percent of values within specified error limits, including within 5%, 10%, 15%, and 20% of the reference value (or within 5 mg/dL [0.28 mmol/L], 10 mg/dL [0.56 mmol/L], 15 mg/dL [0.83 mmol/L], and 20 mg/dL [1.12 mmol/L] of the reference value when the reference value was $<$100 mg/dL [5.55 mmol/L]), and we defined data pair differences exceeding 20% or 20 mg/dL (1.11 mmol/L) as extreme outliers, *3*) modified Bland-Altman analysis comparing the difference between the BGM reading and the reference value as a percentage of the reference value, including average bias, coefficient of variation, 95% limits of agreement, the larger absolute limit boundary, and a modified Bland-Altman plot; and *4*) clinical risk using eight absolute levels in the Surveillance Error Grid (SEG) (12). The SEG is a modern metric for clinical accuracy of BGMs based on risk assessments of BGM errors by diabetes clinicians that assigns a unique risk score to each system-measured data point when compared with a reference value. The SEG specifies the clinical accuracy of a BGM to be portrayed as the percentages of data points falling into prespecified risk zones. This tool can be used to assist regulators and manufacturers to monitor and evaluate BGM performance in their surveillance programs.

RESULTS

Subjects

A total of 1,035 subjects were recruited and enrolled in 2016, of whom 1,032 subjects completed the study (Rainier Clinical Research Center, Inc., 352 subjects; Diablo Clinical Research, Inc., 335 subjects; and AMCR Institute, Inc., 345 subjects). In each study, on average, a BGM was tested on 115 subjects. One enrolled subject dropped out prior to completing the finger-sticks; two subjects' plasma specimens were lost in shipping to the reference laboratory. The sexes, types of diabetes, age races, and ethnicities of subjects are presented in Table 2; however, the study was not intended or powered to study outcomes in any demographic subset. No adverse events occurred. The average number of strip lots per study was 2.1 (SD 1.1 and range 1–4) where evaluable strips could be located. One BGM (TRUEtrack) was excluded from one of the three studies because a recall made strips unavailable.

Reference Instrument Bias

Relative to NIST standards, the mean (SD) bias for each of the five YSI instruments used (in decreasing order of frequency of use) was −0.63% (1.88), −0.24% (1.68), −0.08% (1.66), −0.24% (1.16), and 0.69% (1.68). For the four NIST specimens that were each run on five different YSI instruments, 20 mean biases (one per glucose concentration per instrument) ranged from −1.67% to 1.78%.

Table 2—Demographics

N	1,035
Sex, *N* (%)	
Female	560 (54.1)
Male	471 (45.5)
Not specified	4 (0.4)
Type of diabetes, *N* (%)	
Type 1	370 (35.7)
Type 2	470 (45.4)
Prediabetes	4 (0.4)
Does not have diabetes	187 (18.1)
Not specified	4 (0.4)
Age (years)	
Range	18–87
Mean (SD)	51.5 (16.5)
Race, *N* (%)	
White/Caucasian	853 (82.4)
Black/African American	91 (8.8)
Native American/ Aboriginal	9 (0.9)
Asian	47 (4.5)
Other and not specified	35 (3.4)
Ethnicity, *N* (%)	
Latino/Hispanic	156 (15.1)
Not Latino/Hispanic	879 (84.9)

Performance Relative to Accuracy Standard

Six of the 18 BGMs met the predetermined accuracy standard in all three studies; 5 BGMs met it in two studies; and 3 met it in one study. Four BGMs did not meet the accuracy standard in any of the three studies (Table 3). As mentioned in RESEARCH DESIGN AND METHODS, the protocol defined a compliant data pair as within 15% for BG >100 mg/dL (5.55 mmol/L) or 15 mg/dL (0.83 mmol/L) for BG <100 mg/dL (5.55 mmol/L). Rankings by overall compliant proportion across the three studies (total compliant readings/total readings) coincided with the rankings by number of studies in which the BGM met the accuracy standard. The overall percentage of data pairs that were compliant for every BGM that met the accuracy standard in all three studies was 95% or higher. This overall percentage of compliance is consistent with the minimum necessary percentage for satisfactory performance according to ISO 15197:2013 and FDA 2016 over-the-counter standards. Furthermore, every BGM meeting the accuracy standard on two or fewer of the three studies had an overall compliant proportion of ≤92% (Table 3). By site, the number of BGMs meeting the accuracy standard were 10 of 18, 12 of 18, and 9 of 17 (with one BGM not tested). In terms of overall compliance percentages, the sites were also similar: 88%, 93%, and 88%.

Error Limits and Extreme Outliers

Rankings varied slightly with different tolerances for compliance (±5% or 5 mg/dL [0.28 mmol/L] of the reference value, ±10% [0.56 mmol/L], and ±20% (1.11 mmol/L) (see Supplementary Table 1). Regarding extreme outlier data points, for each of the six BGMs that met the accuracy standard on all three of studies, <2% of readings were >20% from the reference value. For the other BGMs, >2% of readings were >20% from the reference value, with the exception of LifeScan OneTouch Verio, which had 1.3% of readings >20% from the reference value.

Modified Bland-Altman Analysis

The results of the modified Bland-Altman analysis comparing the difference (BGM reading minus reference value) with the reference value were similar to the overall compliance results. The bias of each of the six top-performing BGMs according to the accuracy standard and of the 12 other BGMs ranged, respectively, from −6.0% to 2.4% (with five of these six showing negative bias) and from −10.1% to 5.9% (with 6 of 12 showing negative bias) (Table 4). The larger absolute 95% limit of agreement combines bias with the coefficient of variation. The larger absolute limit boundaries for the six top-performing BGMs and for the 12 other BGMs were, respectively, 11–19% and 19–43%. Although it did not meet the accuracy standard on all three studies, the LifeScan OneTouch Ultra 2 had a larger absolute limit boundary of 19%, which tied the highest value of among the six top-performing BGMs according to the other metrics. The modified Bland-Altman plots are available upon request to the corresponding author.

Clinical Accuracy

The SEG divides BGM reference pairs into eight risk levels, the lowest of which is "no risk." The six top-performing BGMs on other metrics had >97% of readings in the "no risk" category, whereas none of the other BGMs had >97% of readings in this "no risk" category (see Supplementary Table 2).

Excluded Data

Of 1,032 plasma samples, 81 (7.8%) were excluded from analysis—63 for >4% variability between duplicate analyses on the comparator instrument runs, 6 for hemolysis, and 12 for other sample problems (including low sample volume, bubbles in the sample, and autocalibration failure). The remaining 951 reference plasma specimens corresponded with 5,584 BGM readings or 5.9 capillary BGM readings per reference specimen. There were fewer than six BGM-measured specimens per reference reading because 122 capillary readings were not performed or were not evaluable; 114 subjects were to be tested with a TRUEtrack BGM at one site, but the system's strips were recalled just before the third study was scheduled, and no replacement strips could be located. Capillary BG testing was not performed with the recalled strip lot. The 114 subjects were tested with five other BGMs. Furthermore, eight BGM readings of capillary BG levels generated error codes. The proportion of reference values excluded did not differ significantly

Table 3—Results of performance on each of their three studies, including total number of studies meeting the predetermined accuracy standard and total number of data points within protocol limits

			Study 1			Study 2			Study 3			Total number of studies meeting accuracy standard/number of valid trials	Total data points within protocol limits	
Brand	BGM	Test strip	*N*	Percentage compliant†	Met accuracy standard?	*N*	Percentage compliant†	Met accuracy standard?	*N*	Percentage compliant†	Met accuracy standard?		*N*	%
Bayer	Contour Next	Contour Next	98	99	Pass	101	100	Yes	113	100	Yes	3/312	311	100
Roche	Accu-Chek Aviva Plus	Accu-Chek Aviva Plus	97	97	Yes	101	100	Yes	113	98	Yes	3/311	306	98
ARKRAY	Walmart ReliOn Confirm (Micro)	ReliOn Confirm/ Micro	100	96	Yes	114	96	Yes	103	99	Yes	3/317	307	97
AgaMatrix	CVS Advanced	CVS Advanced	101	96	Yes	114	96	Yes	103	98	Yes	3/318	307	97
Abbott Diabetes Care	FreeStyle Lite	FreeStyle Lite	98	92	Yes	101	96	Yes	113	98	Yes	3/312	298	96
Roche	Accu-Chek SmartView	Accu-Chek SmartView	108	98	Yes	106	96	Yes	106	92	Yes	3/320	305	95
ARKRAY	Walmart ReliOn Prime	ReliOn Prime	98	85	No	101	95	Yes	113	96	Yes	2/312	288	92
LifeScan	OneTouch Verio	OneTouch Verio	108	87	No	106	98	Yes	105	91	Yes	2/319	294	92
Prodigy	Prodigy AutoCode	Prodigy No Coding	98	86	No	101	92	Yes	113	93	Yes	2/312	282	90
LifeScan	OneTouch Ultra 2	OneTouch Ultra	97	92	Yes	101	84	No	113	94	Yes	2/311	280	90
Abbott Diabetes Care	Walmart ReliOn Ultima	ReliOn Ultima	107	96	Yes	106	97	Yes	106	75	No	2/319	285	89
Bayer	Contour Classic	Contour	108	95	Yes	106	85	No	106	86	No	1/320	284	89
Omnis Health	Embrace	Embrace No-Code	102	87	No	114	93	Yes	103	84	No	1/319	282	88
HDI/Nipro	TRUEresult	TRUEresult	101	94	Yes	114	83	No	103	86	No	1/318	279	88
HDI/Nipro	TRUEtrack	TRUEtrack	102	83	No	103	80	No	—	—	No*	0/205	167	81
BioSense Medical	Solus V2	Solus	108	56	No	106	84	No	106	89	No	0/320	244	76
Diabetic Supply of Suncoast	Advocate Redi-Code +	Advocate	102	88	No	114	71	No	103	68	No	0/319	241	76
Philosys, Inc.	Gmate SMART	Gmate	108	61	No	106	79	No	106	72	No	0/320	226	71

In studies 1, 2, and 3, each of the 18 BGMs was tested at three different sites. Over the three studies, each BGM was tested once by each site. The number of compliant readings needed to meet the accuracy standard depends on the number of trials. For 100 trials, at least 91 readings must be within 15% or 15 mg/dL (0.83 mmol/L) of the reference value. *No study 3 data for HDI/Nipro TRUEtrack because of test strip recall. †Within 15% of reference value if ≥100 (5.55 mmol/L) or 15 mg/dL (0.83 mmol/L) of reference value if ≤100 mg/dL (5.55 mmol/L).

Table 4—Summary of modified Bland-Altman comparison

BGM system	Valid trials	Bias (%)	Coefficient of variation (%)*	95% limits of agreement†		Larger absolute limit boundary (%)‡
				Lower limit	Upper limit	
Contour Next	312	−1.2	5.3	−11	10	11
Accu-Chek Aviva Plus	311	−3.4	6.3	−15	9	15
Walmart ReliOn Confirm (Micro)	317	2.4	6.8	−10	17	17
CVS Advanced	318	−0.3	7.0	−13	14	14
FreeStyle Lite	312	−6.0	7.4	−19	9	19
Accu-Chek SmartView	320	−5.3	6.5	−17	8	17
Walmart ReliOn Prime	312	0.4	9.5	−17	21	21
OneTouch Verio	319	5.9	6.8	−7	21	21
Prodigy AutoCode	312	1.2	10.3	−17	24	24
OneTouch Ultra 2	311	−3.0	9.3	−19	17	19
Walmart ReliOn Ultima	319	3.0	10.2	−16	26	26
Contour Classic	320	−6.5	9.7	−23	13	23
Embrace	319	0.9	10.0	−17	23	23
TRUEresult	318	−8.9	8.2	−22	7	22
TRUEtrack	205	−6.9	10.6	−24	15	24
Solus V2	320	−10.1	8.1	−23	5	23
Advocate Redi-Code +	319	−9.1	10.5	−26	12	26
Gmate SMART	320	5.7	15.5	−22	43	43

Modified Bland-Altman analysis compares the difference (BGM reading − reference value) with the reference value rather than comparing the difference with the average of the BGM reading and the reference value. The bias is the average difference as a percent of the reference value. A bias of −5.0% means the BGM meter reading is, on average, 5% lower than the reference value. *SD of the difference in the log-transformed measurements, which is essentially the same as the SD of the % difference = (BGM reading − reference value)/(reference value). †95% limits of agreement define the range around the reference value containing 95% of the BGM readings. ‡Limit boundary with the larger absolute value.

between the 6 top-performing BGMs and the 12 other BGMs.

The data analyses included samples with hematocrit ranges that were outside of the product labeling of certain BGMs. However, in an additional analysis, exclusion of 57 specimens with hematocrit outside the narrowest range for any of the 18 BGMs and 19 specimens for which no hematocrit was recorded had no effect on the BGM performance rankings (data not shown).

Sequence Effect

The position in the sequence did not significantly affect the difference from the reference value (P = 0.87). Furthermore, analysis after completion of the study showed that each BGM was equally likely to be tested in each available sequence position.

CONCLUSIONS

We found that of 18 commercially available BGMs, 6 met a predefined accuracy standard on three out of three studies. This accuracy standard was similar to, but more lenient than, those currently used by the FDA. The other metrics of accuracy confirmed the rankings based on meeting the accuracy standard. Therefore, based on our findings it appears that cleared BGMs do not always perform to the level of analytical accuracy that is currently required for clearance.

The six top-performing BGMs according to the accuracy standard also performed the best according to four additional metrics: *1*) overall compliant proportion, *2*) the proportion of extreme outliers (although the LifeScan OneTouch Verio also performed well on this metric), *3*) the greater 95% limit of agreement, and *4*) the proportion in the lowest clinical risk category according to the SEG. Among the 12 lower-ranking BGMs, there was a wide spectrum of overall performance, ranging from meeting the accuracy standard on two, one, or zero out of three trials and demonstrating an overall compliant proportion ranging from 71 to 92%.

In terms of clinical consensus accuracy, the six top-performing BGMs had at least 97% of their data points in the SEG no-risk zone and the other 12 BGMs had <97% of their data points in the SEG no-risk zone. Kovatchev et al. (13) used modeling to calculate that a device with ≤3% errors outside of the SEG no-risk "green" zone would meet the ISO requirements of ≤5% data pairs outside the 15 mg/dL (0.83 mmol/L)/15% standard limits, while higher percentages outside the SEG no-risk zone would indicate noncompliance with the standard. No empirical series to our knowledge has specified a target for clinical accuracy using the SEG, but based on risk zone results of 18 BGMs from this study and a post hoc analysis of these results, we propose that a cutoff for excellent clinical accuracy can be defined as ≥97% of data points in the no-risk zone of the SEG, as Kovatchev et al. had predicted.

Strips were purchased based on availability irrespective of lot number. A defective strip lot could not be ruled out as the cause of poor performance for a given product. The purpose of the study was to ascertain whether there was poor performance of the tested products. The study was not intended to seek out three different lots of any product. The study was also not intended to identify any specific strip lots associated with poor performance of a product.

Many factors affect the accuracy of a BGM, including those related to the test strip and the meter (14,15). Differences in accuracy were not unexpected because

technological factors vary among the BGMs in this study. Although BGMs must now meet criteria similar to the ones we used in this study in order to receive clearance from the FDA to market in the U.S., some currently marketed older BGMs were cleared when accuracy standards were 20% (15 mg/dL [0.83 mmol/L]) per ISO 15197:2003 rather than the current ±15% requirements per the FDA 2016 over-the-counter standards.

The performance of BGMs may diminish over time (i.e., postmarket performance may deteriorate). This decline may be due to scale-up issues, manufacturing errors, changes in components between strip lots, other production issues, or improper shipping. Over time, the measured analytical accuracy might no longer represent the sponsor's initial accuracy data that were submitted to the FDA. Such factors might account for our findings.

To assess whether there was significant year-to-year turnover in market share of the most widely purchased BGMs, we compared the Medicare mail-order shares distribution of the top BGMs purchased between quarter 4 of 2013 (9), when this surveillance program was first planned (16), and quarter 2 of 2016 (17) when we began our study. According to that database, the mail-order shares for the 18 BGMs that we tested changed from 90.1% to 84.3% over that 2.5-year period, which indicated only a small annual turnover.

The performance levels in this surveillance protocol represent how each BGM product functioned in our research study carried out by trained medical professionals. This performance cannot necessarily be extrapolated to use by patients. The total number of times that a BGM met the protocol's accuracy standard as tested by health care professionals on a particular set of strips and meters at a specific time does not mean that a patient or other user can expect any particular performance from the product other than what the manufacturer claims. Product performance can change over time. The authors make no claims, endorsements, or predictions for future performance of the tested products.

Strengths of this study include the large number of subjects tested (1,032), the large number of data pairs evaluated for agreement (5,584), the large number of FDA-cleared BGMs tested (18 systems tested three times each), and the consistency of the outcomes achieved by several evaluation methods (e.g., number of studies meeting the accuracy standard, overall compliant proportion, frequency of extreme outliers, modified Bland-Altman analysis, and clinical accuracy using the SEG). To our knowledge, this is the largest accuracy study of FDA-cleared BGMs using a consensus protocol created with input from the FDA ever reported in the literature. All strips and monitors were purchased from commercial suppliers without the manufacturers' knowledge to avoid positive bias that could occur if a manufacturer were to have an opportunity to submit their best performing strips or monitors for testing. Also, the protocol was developed by an impartial expert panel. Testing performed by health care professionals tends to lead to more accurate results than when subjects test themselves (18), which could lead to a higher level of accuracy in this study compared with other studies where subjects self-test. Finally, this study was triple blinded, which eliminated the possibility of systematic bias based on BGM brand.

A limitation of this study is the exclusion of 81 out of 1,032 (7.8%) of the reference samples. Another limitation is a mean downward drift of 0.12% between the time points of the capillary tests, 0.52% from the first to the sixth BGM tested. However, this decrease was not statistically significant and could not have biased findings in favor or against a specific BGM, since the testing sequence was randomized. Yet another limitation is that products are frequently replaced by newer models and some of the products tested may not remain on the market for a prolonged period of time in the future. The BGMs tested in this study are all intended only for outpatient self-monitoring. A similar study of prescription point-of-care BGMs used in hospitals and nursing homes could be performed in the future.

In conclusion, 6 of the 18 best-selling personal BGMs met a protocol-specified accuracy standard similar to current ISO and FDA standards on three of three studies. These same six meters ranked highest according to four other metrics. Since patients depend on their BGMs for day-to-day management, lack of accuracy may put patients at risk for both hypoglycemia and hyperglycemia. We believe that this study points out the varying degrees to which commonly used BGMs do or do not give accurate information. We hope that this study will provide objective and validated information for patients, health care professionals, and payers to make informed product selection. We also hope that this study will provide important information that will lead regulators to consider introducing a mechanism to evaluate postmarket performance of these types of analytical products.

Acknowledgments. The authors thank the following persons for their contributions to conducting this study: Laura Bedolla (AMCR Institute, Inc.), Pam Martin (Rainier Clinical Research Center, Inc.), and Catherine Morimoto (Diablo Clinical Research) for recruiting subjects and performing capillary BG measurements; Dr. Kristin Castorino (William Sansum Diabetes Center) for overseeing performance of reference glucose measurements; and Megha Shah (clinical research professional) for oversight of the clinical sites. The authors also thank Mike Jarrett (QuesGen Systems, Inc., Burlingame, CA) for assistance with data analysis, Dan Shilstone (Diabetes Technology Society, Burlingame, CA) for assistance with data presentation, Brett McGreevy (Diabetes Technology Society) for managerial oversight of the project, and Annamarie Sucher (Diabetes Technology Society) for expert editorial assistance.

Duality of Interest. This study was supported by a grant from Abbott Diabetes Care. D.C.K. is a consultant for Ascensia, AstraZeneca, EOFlow, Intarcia, Lifecare, Novo Nordisk, and Voluntis; has received research funding from Diasome, Lexicon, and Novo Nordisk; and is an employee of DTS. J.L.P. is a Bayer retiree and a consultant for DTS. B.P.K. has received grant/research support from Dexcom, Roche Diabetes Care, Sanofi, Senseonics, and Tandem Diabetes Care; is on the advisory board, is a consultant, and is on the speaker's bureau for Dexcom, Sanofi, and Senseonics; is a stock shareholder for TypeZero Technologies; and has patent royalties managed by the University of Virginia Licensing and Ventures group from LifeScan, Animas, and Sanofi. D.K. is a medical advisor to Glooko and Vicentra and is creator of www.diabetestravel.org and www.excarbs.com. William Sansum Diabetes Center has received research funding from Abbott Diabetes Care, Dexcom, Sanofi, Novo Nordisk, and Lilly. R.L.B. received research grant support from Abbott Diabetes Care, Roche, Bayer, Senseonics, Dexcom, and Medtronic. M.C. has received research funding from Abbott Diabetes Care, Bayer, Dexcom, Insulet, Medtronic, and Senseonics. T.S.B. has received research support from Abbott Diabetes Care, Ambra, Ascensia, BD, Boehringer Ingelheim, Calibra, Companion Medical, Dexcom, Elcelyx, GlySens, Janssen, Lexicon, Lilly, Medtronic, Novo Nordisk, Sanofi, Senseonics, Versartis, and Xeris; consulting honoraria from AstraZeneca, Ascensia, BD, Calibra, Lilly, Medtronic, Novo Nordisk, and Sanofi; and speaking honoraria from Abbott Diabetes Care, Insulet, Medtronic, Lilly, Novo Nordisk, and Sanofi. J.H.N. has accepted honoraria and travel expenses for professional speaking, consulting, and participation in scientific advisory boards for Abbott Laboratories and Roche Diagnostics. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. D.C.K. wrote the manuscript and organized the study. J.L.P. wrote the manuscript and contributed to the discussion. B.P.K. reviewed the manuscript and advised about the statistics. D.K. reviewed and edited the manuscript. W.C.B. researched data at the reference laboratory. R.L.B. researched data at a clinical site. M.C. researched data at a clinical site. T.S.B. researched data at a clinical site. J.H.N. contributed to the discussion. M.A.K. wrote the manuscript and advised about the statistics. D.C.K. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. The summary results were first reported on the DTS website in preparation for a presentation at the U.S. Congressional Diabetes Caucus, Washington, DC, 23 August 2017 (https://www.diabetestechnology.org/surveillance.shtml).

References

1. Choudhary P, Genovese S, Reach G. Blood glucose pattern management in diabetes: creating order from disorder. J Diabetes Sci Technol 2013;7:1575–1584

2. 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: an in silico study. J Diabetes Sci Technol 2010;4:562–570

3. International Organization for Standardization. *ISO 15197:2013 In Vitro Diagnostic Test Systems – Requirements for Blood-Glucose Monitoring Systems for Self-Testing in Managing Diabetes Mellitus*, Geneva, Switzerland, 2013

4. U.S. Department of Health and Human Services, U.S. Food and Drug Administration. Self-monitoring blood glucose test systems for over-the-counter use: guidance for industry and Food and Drug Administration staff [Internet], 2016. Available from https://www.fda.gov/downloads/ucm380327.pdf. Accessed 24 May 2018

5. Klonoff DC, Prahalad P. Performance of cleared blood glucose monitors. J Diabetes Sci Technol 2015;9:895–910

6. Freckmann G, Baumstark A, Pleus S. Do the new FDA guidance documents help improving performance of blood glucose monitoring systems compared with ISO 15197? J Diabetes Sci Technol 2017;11:1240–1246

7. Ekhlaspour L, Mondesir D, Lautsch N, et al. Comparative accuracy of 17 point-of-care glucose meters. J Diabetes Sci Technol 2017;11:558–566

8. 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;10:697–707

9. IMS Health Information Service. National Prescription Audit. IMS therapeutic class 40440 diagnostic aids blood glucose test, ZA Diag strips, sticks, tape, for the period 12 months ending 07/2015 (database)

10. Murrin S. *Memorandum Report: Medicare Market Shares of Mail Oder Diabetes Test Strips 3-6 Months After the Start of the National Mail Order Program, OEI-04-13-00682*. Department of Health and Human Services, Office of Inspector General, Washington, DC, 2014

11. Fein AJ. *2014-15 Economic Report on Retail, Mail, and Specialty Pharmacies*. Drug Channels Institute, 2015

12. Klonoff DC, Lias C, Vigersky R, et al.; Error Grid Panel. The surveillance error grid. J Diabetes Sci Technol 2014;8:658–672

13. Kovatchev BP, Wakeman CA, Breton MD, et al. Computing the surveillance error grid analysis: procedure and examples. J Diabetes Sci Technol 2014;8:673–684

14. Tonyushkina K, Nichols JH. Glucose meters: a review of technical challenges to obtaining accurate results. J Diabetes Sci Technol 2009;3:971–980

15. Schmid C, Haug C, Heinemann L, Freckmann G. System accuracy of blood glucose monitoring systems: impact of use by patients and ambient conditions. Diabetes Technol Ther 2013;15:889–896

16. Diabetes Technology Society. Diabetes Technology Society outlines potential post-clearance surveillance program to address inaccuracy of blood glucose monitors [article online], 2013. Available from https://www.prnewswire.com/news-releases/diabetes-technology-society-outlines-potential-post-clearance-surveillance-program-to-address-inaccuracy-of-blood-glucose-monitors-223628961.html. Accessed 18 December 2017

17. U.S. Department of Health and Human Services, Office of Inspector General. Medicare market shares of mail order diabetes test strips from April to June 2016 [article online], 2016. Available from https://oig.hhs.gov/oei/reports/oei-04-16-00470.pdf. Accessed 18 December 2017

18. Heinemann L. Quality of glucose measurement with blood glucose meters at the point-of-care: relevance of interfering factors. Diabetes Technol Ther 2010;12:847–857