Supplementary Data

Supplementary Table S1. Sources for Best Practice Quality Guidelines or Recommendations (Modified from Mahoney and Ellison[1,2])

- ISO 15197: The International Standards Organization[6] and recent new draft[7]
- FDA: U.S. Food and Drug Administration (FDA 2006)[34]
- IFCC: International Federation of Clinical Chemistry[35]
- SKUP: Scandinavian Evaluation of Laboratory Equipment for Primary Health Care[36]
- MHRA: UK Medicines and Healthcare Products Regulatory Agency[37]
- National Standard of the People's Republic of China[38]
- STARD: Standards for Reporting Diagnostic Accuracy[39]
- CLSI: Clinical and Laboratory Standards Institute[40,41]
- TNO: Netherlands Organization for Applied Scientific Research[42]

Supplementary Table S2. Aspects of Accuracy Evaluation Given in Guidelines or Recommendations

- appropriate blood specimens
- blood sample collection method
- blood sample handling
- protocols
- use of defined blood glucose meter systems by trained operators in accordance with the manufacturer's instructions for use
- the reference method
- statistically validated presentation of results
- assessment against acceptance criteria
- inclusion of appropriate details and sufficient discussion of results

Supplementary Table S3. Publications Examined

Studies assessing	*Reference (year)*
Accuracy of BG systems	
1. Accuracy of four BG monitoring systems	Schwartz et al.[12] (2008)
2. A glucose meter accuracy and precision comparison: the FreeStyle Flash versus the Accu-Chek Advantage, Accu-Chek Compact Plus, Ascensia Contour, and the BD Logic	Thomas et al.[13] (2008)
3. Accuracy study of BG monitoring systems	Bergenstal[14] (2011)
4. System accuracy evaluation of 27 BG monitoring systems according to DIN EN ISO 15197	Freckmann et al.[15] (2010)
5. Clinical study measuring the accuracy of the mylife Pura BG meter in the management of intensive insulin therapy in patients with type 1 diabetes	Mentis et al.[16] (2010)
6. Can glucose meters meet tighter accuracy requirements?	Ng and Lock[17] (2010)
7. Individuals achieve more accurate results with meters that are codeless and employ dynamic electrochemistry	Rao et al.[18] (2010)
8. Accuracy and precision evaluation of seven self-monitoring BG systems	Kuo et al.[19] (2011)
9. Accuracy and precision of glucose monitoring are relevant to treatment decision-making and clinical outcome in hospitalized patients with diabetes to treatment decision-making and clinical outcome in hospitalized patients with diabetes	Voulgari et al.[20] (2011)
10. Accuracy evaluation of five BG monitoring systems obtained from the pharmacy: a European multicenter study with 453 subjects	Tack et al.[21] (2012)
Accuracy of BG systems and influence of interferences	
1. Standardized evaluation of nine instruments for self-monitoring of BG	Kristensen et al.[22] (2008)
2. Evaluation of the analytical specificity and clinical application of a new generation hospital-based glucose meter in a dialysis setting	Bewley et al.[23] (2009)
3. An evaluation of the analytical performance of a new-generation hospital-based glucose meter and an assessment of its clinical reliability in a neonatal care unit	Thomas et al.[24] (2009)
4. Glucose meters: evaluation of the new formulation measuring strips from Roche (Accu-Chek) and Abbott (MediSense)	Dimeski et al.[25] (2010)
5. Can one POC glucose meter be used for all pediatric and adult hospital patients? Evaluation of three meters, including recently modified test strips	Warner et al.[26] (2011)
Influence of interferences	
1. Performance of BG measurement systems influenced by interfering substances	Pfützner et al.[27] (2009)
2. Dynamic electrochemistry corrects for hematocrit interference on BG determinations with patient self-measurement devices.	Musholt et al.[28] (2011)
3. How accurate are BG meters used for patient self-testing?	O'Kane et al.[29] (2011)
4. Effect of ambient temperature on analytical performance of self-monitoring BG systems	Nerhus et al.[30] (2011)
5. Interferents in glucose determination do not influence the hospital POC glucose meter StatStrip in accuracy and precision of BG measurement.	Schöndorf et al.[31] (2011)

BG, blood glucose; POC, point-of-care.

Supplementary Table S4. Criteria for Assessing Quality of Glucose Meter System Accuracy Evaluation Studies

1. Study design, independence, and impartiality:

   A prospective study design, which allows investigators to determine the appropriateness of blood samples and control a range of potentially variable protocol factors, should be used.[2]

   Studies should be perceived as unbiased and independent and performed by an objective third party. They should be conducted at external sites such as outpatient clinics or hospital settings to mimic real-life use and eliminate any manufacturer bias.

2. Study population:

   Patients and samples included in the study should satisfy appropriate inclusion and exclusion criteria,[2] which should be population dependent and take into account limitations in the BG system labeling. Possible contraindications and exclusion criteria include patient contraindications, which are:

   - device-independent (e.g., specimens from patients in shock or dehydration)
   - device-dependent (e.g., specimens containing inappropriate oxygen levels, extremes of hematocrit, or known interfering substances)

   Interference studies on device-dependent factors such as for hematocrit, maltose, temperature, and reducing substances are treated separately and assessed against ISO 15197[7] and CLSI[8] recommendations for interference testing.

3. Glucose meter calibration and units:

   Glucose monitors that test whole blood and display "plasma-equivalent" data using a 1.11 conversion factor are recommended.[35,42] This assumes a single hematocrit measurement for conversion, and a better possible approach is to use a formula using the patient's actual hematocrit level.

4. Number of donors, number of samples, and spread of results:

   Using whole blood samples from a minimum of 40 different donors or 40 residual blood samples from 40 different donors has been advocated.[40] However, this has been reported to be insufficient,[3] gives an unreasonable high chance of rejecting a good system, and is underpowered for rejecting poor performers. Samples from more donors will improve the confidence of the estimates and may be important in certain study designs.[41] ISO 15197 guidelines recommend use of at least 100 fresh capillary samples and at least 200 data points to reach conclusions.[6]

   There should be a sufficient spread of results spanning the analytical range with appropriate percentages of results within specific concentration intervals. An appropriate distribution is specified in ISO 15197. Pooling of whole blood samples into a single test sample is not recommended.[2]

   Capillary samples with very high/low glucose concentrations can be provided by using appropriately modified samples prepared by validated processes substituted for "fresh" samples.

5. BG meter operators:

   Operators must use the meter system according to the manufacturer's instructions. They should be familiar with the system, and have knowledge of and be trained in how to avoid pre-analytical, analytical, and post-analytical errors. A combination of healthcare professional or patient operators may be helpful depending on the design and intention of the research.[34,36,43]

   Clinical and other testing personnel who perform the glucose tests should be trained on the device limitations, manufacturer's instructions for use, safety practices, and the test protocol.[2]

   Operators testing the reference sample should also be trained in the correct use of the reference assay and equipment.

6. Blood sample type and comparing "like with like" samples

   Appropriate comparisons of "like" specimens, such as capillary versus capillary comparisons, should be made. Capillary/venous/plasma/whole blood comparisons will exhibit natural differences and inappropriately alter accuracy conclusions. The glucose concentration in capillary and venous blood should not be assumed to be equivalent. Comparisons may show differences of about 2%, but there can be up to 30% differences in the postprandial state.[44]

   Only fresh human whole blood sample types that are listed as appropriate for the glucose system should be used.[6] Most meters have been calibrated to test capillary whole blood; others have the capability to test capillary, venous, arterial, or neonatal blood samples.

   Additives to the test samples (e.g., anticoagulants) are permitted only if specifically stated in the device labeling.[6]

   Unless the subject of specific interference studies, the hematocrit, the partial pressure of oxygen in the blood sample, and other conditions must be within the stated limitations of the device. Any exclusion criteria for samples should be explained and based on the instructions for use.

   Artificial materials and manufacturer's control solutions are not recommended for assessing the accuracy of glucose monitors.[6]

   Control of sampling time is important as after a carbohydrate load BG can change rapidly at a sampling site.[45] There should be appropriate minimal delay in analysis and post-collection control of sample handling time is important because glycolysis can cause rapid glycemic change dependent on the hematocrit.[46] If either of these is not controlled, differences in the data can be due to glucose concentration differences in the comparative samples instead of differences between the two methods.[1]

   Sample collection procedures and application should be explained, and ideally samples should be applied directly to strips to mimic actual routine use.

7. Number of strip/reagent "lots" and meter system testing:

   Each sample should ideally be tested in duplicate according to the manufacturer's instructions for use.[6] Single testing has been indicated as appropriate, avoiding complexity and pain/inconvenience to patients.[3]

   Recommendations for system handling and storage must be followed.

(*continued*)

Test strips and meters that have been stored improperly or whose storage conditions are not known should not be included in the evaluation.

The ambient test conditions (e.g., temperature and humidity) must be within the device specifications and information on strip lots, examination time, date of expiry, and meter serial numbers should be provided.

A. Reagent strips:

Conclusions should be based on studies providing performance data on more than one "lot" of strips.[6] This indicates the robustness of accuracy conclusions and should be representative of product intended for sale.

At least 200 units (strips) from at least 10 vials should be used.[6] The number of lots used should be stated. Use of only one lot is allowed, but use of several lots, ideally three different lots, indicates robustness.

B. Meters:

More than one BG meter per subject may be needed to minimize time between duplicates samples. Samples should be measured with two different BG meters, which should be rotated during the protocol and used with equal frequency.

8. Reference method:

Meter results should be compared against results generated by the reference method specified by the manufacturers.[6] Five percent differences are common if inappropriate reference methods are used, and laboratory reference methods for blood glucose can have a total error of up to 10%.[47] Information should be supplied on the reference method, if reference samples were analyzed in at least duplicate, imprecision, quality assurance, and its traceability to higher reference methods.

The reference method must be traceable to materials or methods of higher order.[9] For glucose testing this means the reference method must be capable of testing glucose in human plasma, serum, or deproteinized whole blood samples. Glucose results determined by blood gas instruments are not generally acceptable as reference results because certified reference materials for glucose in whole blood are not yet available.[2]

The reference method should be checked and monitored for stability using appropriate quality control materials and procedures. Reference sample analysis should be performed in at least duplicate and checked for differences of greater than 4% (or 0.22 mmol/L [4 mg/dL]) at 5.6 mmol/L (100 mg/dL).[6]

The trueness of the reference method should be checked with National Institute of Standards and Technology reference materials or other traceable materials.

Ideally the enzyme of the reference method should be comparable to the one used by the BG system to avoid problems of specificity.[2] Enzymes including glucose oxidase, hexokinase, and glucose dehydrogenase are used in reference laboratory analyzers, which commonly include YSI or Hitachi instruments.

The bias for the reference method should be 2.2%, analytical imprecision 2.9%, and total error 6.9%.[48] Based on imprecision, not exceeding one-half of the within-individual biological coefficient of variation, an imprecision of $< 2.2\%$ and a 0% bias has been suggested as a target.[48]

9. Methodology

a. Test in controlled temperature and humidity, within manufacturers' specifications, generally $23 \pm 5°C$.[6]
b. Operators must be trained, be familiar with the proper operation of the BG system, and operate the system according to the manufacturer's instructions for use.
c. Follow CLSI guidelines for safe skin puncture, blood collection, and handling. Ensure the skin site is clean and dry prior to skin puncture.
d. For capillary blood, first obtain a fingerstick puncture for the reference method without excessive pressure or squeezing. If there is sufficient volume, a portion can be applied to the BG system following the manufacturer's instructions for use.[2] Blood from multiple fingertip punctures from different fingers may be necessary. Perform at least two BG system tests within 5 min of each other.[6,36] A blood sample can then be obtained for a second reference test and a hematocrit level, although it has been proposed that if precautions are taken to ensure the BG does not change, one sample is sufficient[3] and avoids duplicate sampling, which may not be appropriate for all patients.
e. For venous (or arterial) blood a "split sample" design should be used that tests both methods with a portion of the same test sample. Gently mix the blood, obtain a portion for the reference method test, test the meter in duplicate, and finally obtain a sample for the second reference method test.
f. Within 5 min of the two meter tests, the reference samples should be centrifuged, and the plasma should be removed. The glucose concentration from a reference sample must be tested with the reference method within an adequate time[3] of the meter tests (e.g., 60 min with a YSI reference[2]) or stabilized (e.g., by a validated deproteinization method) for later measurements.
g. Test the blood for percentage hematocrit and verify it is within the meter manufacturer's acceptable range. If outside these limits exclude the results.
h. Repeat steps (c)–(g) until 99 or more acceptable specimens have been collected. These should cover a wide range of glucose values, ideally with at least defined percentages within specified concentration ranges. Lowering or elevating the glucose of additional samples, via glycolysis or spiking, may be performed using validated methods.[6]

10. Sample stability evaluation:

Compare the values from the two reference samples and verify they are within 4% or 0.22 mmol/L (4 mg/dL) or else exclude the results.[6,36] Calculate the average of the two reference test results.

A major role of the two reference tests is to control preanalytical steps and the quality of sample preparation. They (a) confirm the response of the reference method is reproducible and has not drifted, (b) the sample BG has remained relatively constant, and (c) any time delay after the first reference test has not markedly affected the meter results.[2]

11. Statistical analysis and "outlier" results:

(*continued*)

Compare individual meter results to the average of the two reference method tests. Differences are calculated as an arithmetic difference when the reference glucose average result is <4.2 mmol/L (<75 mg/dL) or a percentage difference for glucose ≥ 4.2 mmol/L (≥ 75 mg/dL).[6]

An analysis of "outlier results" should be performed.[41] Sufficient information should be supplied about "outlier" results, their discrepancies, and why they were omitted.

12. Presentation of results and acceptance criteria for accuracy:

Results should be analyzed for accuracy and displayed in an appropriate manner. Ideally the accuracy claim should be representative of "all" lots of strips and actual lots made available for sale.

Accuracy assessment is best described in terms of:

a. A simple difference plot of the difference between individual results from meters against the mean of specific reference values plotted as the dependent variable.
b. Tables displaying the degree of meter difference compared to the reference method.

The first table should deal with reference glucose values <4.2 mmol/L (<75 mg/dL) and provide the number of meter samples (%) within ± 0.28 mmol/L (5 mg/dL), ± 0.56 mmol/L (10 mg/dL), and ± 0.83 mmol/L (15 mg/dL) of the reference method.

The second table should deal with reference values ≥ 4.2 mmol/L (≥ 75 mg/dL) and provide the number of samples (%) within $\pm 5\%, \pm 10\%, \pm 15\%$, and $\pm 20\%$ of the reference method.

c. A summary of the results (including any results identified as statistical outliers) identified as acceptable <4.2 mmol/L (<75 mg/dL) (within ± 0.28 mmol/L [5 mg/dL]) added to the number of acceptable results ≥ 4.2 mmol/L (≥ 75 mg/dL) (within $\pm 5\%, \pm 10\%, \pm 15\%$, and $\pm 20\%$).[2,6,34]

The new draft of ISO 15197[7] defines accuracy by the percentage of results within ± 15 mg/dL (0.83 mmol/L) for glucose values <100 mg/dL (5.6 mmol/L), and within $\pm 15\%$ for glucose concentrations ≥ 100 mg/dL.[7]

d. A clinical accuracy assessment such as by Parkes or consensus error grid analysis.[49]

Information should be provided on the total number of samples analyzed, the range and spread of glucose concentrations, regression analysis, and line of identity. A summary of statistics with confidence limits, a summary of any outliers, and references for statistical analysis procedures can also be included.

13. Acceptance criteria:

The current minimum acceptance criterion is that 95% of the glucose meter results are accurate. At glucose concentrations of <4.2 mmol/L (<75 mg/dL), clinical accuracy is defined as the percentage of results within ± 0.83 mmol/L (15 mg/dL), and at glucose concentrations ≥ 4.2 mmol/L, 95% of results should be within $\pm 20\%$ of the reference method.[2,6,34,36] From the ideally 200 results (including any results identified as statistical outliers) the number of acceptable results (<4.2 mmol/L added to the number of acceptable results ≥ 4.2 mmol/L) should be stated.

The new draft of ISO 15197[7] defines accuracy by the percentage of results within ± 15 mg/dL (0.83 mmol/L) for glucose values <100 mg/dL (<5.6 mmol/L) and within $\pm 15\%$ for glucose concentrations ≥ 100 mg/dL.[7] It includes protocols for both laboratory personnel and patients as the intended users and supplies the same acceptance criteria for both.

It is important to remember the total error criterion includes the error associated with the reference method. In addition, the number of samples, the glucose distribution of samples, random patient interferences, and residual protocol error can all influence the outcome.[2]

Acceptable deviation from reference results is an area of considerable conflict between many regional or national guidelines. Guidelines and minimum standards adopted, such as those from CLSI and ISO, should be based on three considerations: (a) the weight of expert medical opinion, (b) the state of art of currently available technology, and (c) the effectiveness of current BG systems as demonstrated in clinical outcome studies using state of the art monitoring systems.[2]

14. Discussion, details provided, and concordance with ISO standard 15197 and other relevant guidelines/recommendations:

Discussion should explain the results, protocol deviations, observations, and limitations. The meter's performance as well as information regarding the bias and imprecision of the reference method should be included.[2]

Full details should be provided to establish if the study was undertaken appropriately in accordance with ISO standard 15197 and other relevant guidelines/recommendations.

Sufficient information should be provided in publications to verify if the study was of appropriate design and the conclusions were justified and correct. ISO standard 15197, for example, contains a full list of requirements and information, which may not be provided in all summaries of studies. This, for example, includes diverse protocol information such as the number of meters, number and details of lots of test strips, dates of expiry, temperature range, exclusion criteria based on the instructions for use, serial numbers of meters, etc.[6]

BG, blood glucose.

Supplementary Table S5. Evaluation of Study Designs Using the Accuracy Checklist

	Yes		*Partial*		*No*	
Accuracy checklist	n	%	n	%	n	%
Reference method	4	27	4	27	7	47
Comparing "like with like" samples	9	60	1	7	5	33
Number of samples	7	47	7	47	1	7
Spread of glucose concentrations	1	7	4	27	10	67
Accuracy criteria	3	20	12	80	0	0
Number of strip "lots"	6	40	9	60	0	0
Full details provided	1	7	5	33	9	60
Independency	7	47	0	0	8	53
Concordance with ISO 15197	1	7	8	53	6	40

Supplementary Table S6. Evaluation of Study Designs Using the Interference Checklist

	Yes		*Partial*		*No*	
Interference checklist	n	%	n	%	n	%
Reference method	3	30	4	40	3	30
Use of appropriate samples	5	50	5	50	0	0
Investigation at different glucose concentrations	8	80	1	10	1	10
Number of interference levels tested/rationale/prep	5	50	5	50	0	0
Details of meter performance	6	60	1	10	3	30
Number of replicates	5	50	4	40	1	10
Presentation of results	4	40	2	20	4	40
Interpretation of results	4	40	6	60	0	0
Full details provided	2	20	6	60	2	20
Independency	8	80	0	0	2	20
Concordance with CLSI EP7 and ISO/DIS 15197:2011	1	10	7	70	2	20

Supplementary References

34. Food and Drug Administration: Total Product Life Cycle for Portable Invasive Blood Glucose Monitoring Systems. Draft Guidance for Industry and FDA Staff. Rockville, MD: Food and Drug Administration, 2006.
35. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, Külpmann WR, Larsson L, Lewenstam A, Maas AH, Mager G, Naskalski JW, Okorodudu AO: Approved IFCC recommendation on reporting results for blood glucose (abbreviated). Clin Chem 2005;51:1573–1576.
36. Scandinavian Evaluation of Laboratory Equipment for Primary Health Care: Report on One-Touch Ultra. Publication number SKUP/2005/39. www.SKUP.nu (accessed September 8, 2012).
37. Medicines and Healthcare Products Regulatory Agency: Point of Care Testing—Blood Glucose Meters. London: Medicines and Healthcare Products Regulatory Agency, 2005.
38. GB/T 19634-2005 In Vitro Diagnostic Test Systems—General Technical Requirements for Blood-Glucose Monitoring Systems for Self-Testing. Publication number GB/T 19634-2005. Beijing: National Standard of the People's Republic of China, 2005.
39. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC: Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Fam Pract 2004; 21:4–10.
40. Sacks DB, Bernhardt P, Dunka LJ Jr., Goldstein DE, Hortin GL, Mueller P: Point-of-Care Blood Glucose Testing in Acute and Chronic Care Facilities; Approved Guideline—Second Edition. NCCLS document C30-A2. Wayne, PA: NCCLS, 2002.
41. Krouwer JS, Tholen DW, Garber CC, Goldschmidt HMJ, Kroll MH, Linnet K, Meier K, Robinowitz M, Kennedy JW: Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Second Edition. NCCLS document EP9-A2. Wayne, PA: NCCLS, 2002.
42. Post H, van Keulen GJ, Muller WJ, Miedema K: TNO Quality Guideline PG/TG/2001.045: Portable In-Vitro Blood Monitor Systems for (Self)-Monitoring—Blood Glucose Monitors—Particular Requirements and Test Methods. Leiden: Netherlands Organisation for Applied Scientific Research, 2001.
43. Kristensen GB, Nerhus K, Thue G, Sandberg S: Standardized evaluation of instruments for self-monitoring of blood glucose by patients and a technologist. Clin Chem 2004; 50:1068–1071.
44. Arabadjief D, Nichols JH: Assessing glucose meter accuracy. Curr Med Res Opin 2006;22:2167–2174.
45. Ellison JM, Stegmann JM, Colner SL, Michael RH, Sharma MK, Ervin KR, Horwitz DL: Rapid changes in postprandial blood glucose produce concentration differences at finger,

forearm, and thigh sampling sites. Diabetes Care 2002;25: 961–964.
46. Sidebottom RA, Williams PR, Kanarek KS: Glucose determinations in plasma and serum: potential error related to increased hematocrit. Clin Chem 1982;28:190–192.
47. Ehrmeyer SS, Laessig RH, Leinweber JE, Oryall JJ: 1990 Medicare/CLIA final rules for proficiency testing: minimum intralaboratory performance characteristics (CV and bias) needed to pass. Clin Chem 1990;36:1736–1740.
48. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS, Lernmark A, Metzger BE, Nathan DM: Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2011;57:6.e1–e47.
49. Parkes JL, Slatin SL, Pardo S, Ginsberg BH: A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. Diabetes Care 2000; 23:1143–1148.