scientific reports

Check for updates

OPEN

Evaluation of the accuracy, precision, and agreement of a glucometer compared to the standard laboratory test in diabetic and non-diabetic patients

Hamid Sharif-Nia[1,2], Hosein Mokhtari[3,4], Jason W. Osborne[5], Shahriar Shafaei[6] & Maryam Soltanzade[7]✉

Blood glucose monitoring is vital for diabetes management, with nurses playing a key role in glycemic monitoring and patient education. Despite their widespread use, the precision, accuracy and agreement of fingerprick glucometers with laboratory standards remain a significant concern. Discrepancies arise from multiple sources such as differences in sample type (whole blood vs. serum) and enzymatic technologies, potentially leading to clinical errors. Current international standards like ISO 15,197 have been insufficient in ensuring reliable device performance. Therefore, we need more rigorous benchmarking standards to assess their safety and reliability. This study aimed to evaluate the analytical and clinical accuracy, precision, and agreement of the glucosureSTAR point-of-care (POCT) glucometer against the standard laboratory method in diabetic and non-diabetic patients, following the ISO 15197:2013 standards. A cross-sectional study was conducted on 589 participants (136 diabetic and 453 non-diabetic). Paired blood glucose measurements were taken using the glucosureSTAR glucometer (glucose oxidase method) and a Roche Hitachi 917 laboratory analyzer. Analytical accuracy was assessed against ISO 15197:2013 criteria (±15 mg/dL for values < 100 mg/dL, ±15% for values ≥ 100 mg/dL). Clinical accuracy was evaluated using Parkes and Clarke Error Grid Analyses. Agreement was statistically analyzed using Bland-Altman plots, Passing-Bablok regression, and the Concordance Correlation Coefficient (CCC). Precision was tested via the coefficient of variation from 20 repeated measurements on three samples. Diagnostic performance was assessed using ROC curve analysis. The analytical accuracy of the glucometer marginally met the ISO 15197:2013 criteria (95.1% of results within ±15 mg/dL or ±15%). While clinical accuracy was acceptable (100% of points in Parkes error grid zones A/B), significant systematic error was observed. The device consistently underestimated blood glucose (mean bias: -3.3 mg/dL; 95% LoA: -38.7 to 32.1 mg/dL) with poor absolute agreement (CCC = 0.886). Precision failed ISO requirements at medium and high concentrations. ROC analysis revealed suboptimal diagnostic performance for detecting diabetes (AUC = 0.816; sensitivity = 62.18%, specificity = 93.26% at > 124 mg/dL cut-off). In conclusion, while the glucosureSTAR glucometer demonstrated acceptable clinical risk assessment with 100% of readings in clinically acceptable error grid zones, it showed significant limitations in analytical performance. The device exhibited systematic underestimation, wide variability, and failed to meet precision requirements at medium and high glucose concentrations. With 95.1% of the data points falling within the acceptable range by ISO15197:2013, and suboptimal diagnostic accuracy (sensitivity: 62.18%), this device should not be considered suitable for diagnostic purposes or therapy adjustment. It may be used with caution for trend monitoring in known diabetic patients where rapid results are prioritized over absolute accuracy but cannot replace standard laboratory methods when precise measurement is critical.

Keywords Blood glucose self-monitoring, Glucometers, Point-of-care testing, Diabetes mellitus, ISO 15197:2013, Analytical accuracy, Clinical agreement, Diagnostic performance

[1]Psychosomatic Research Center , Mazandaran University of Medical Sciences, Sari, Iran. [2]Department of Nursing Amol Faculty of Nursing and Midwifery, Mazandaran University of Medical Sciences, Sari, Iran. [3]Immunogenetics Research Center, Mazandaran University of Medical Sciences, Sari, Iran. [4]Department of Paramedicine, Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, Iran. [5]Department of Statistics, Miami University, Oxford, OH, USA. [6]Retired, Department of Pathology, Faculty of Medicine, Babol University of Medical Sciences, Babol, Iran. [7]Student Research Committee, Mazandaran University of Medical Sciences, Sari, Iran. ✉email: mary.snz2000@gmail.com

Blood glucose monitoring has been recognized as one of the principal methods for diabetes management and treatment in recent decades[1]. Consequently, diverse business models have been developed with the aim of expanding the variety of blood glucose measurement tools utilizing various technologies[2,3]. The primary objective of these advancements is to facilitate the process of monitoring blood glucose levels, enhance the accuracy and efficiency of monitoring, and ultimately prevent healthcare costs and disabling complications associated with diabetes[2,4]. A wide spectrum of equipment, ranging from laboratory devices to advanced statistical algorithms, is capable of predicting blood glucose levels and alerting to the occurrence of hypoglycemic and hyperglycemic conditions[5–7]. Nevertheless, fingertip glucometers continue to be widely used as common and trusted tools in healthcare settings by medical staff and patients alike[3,8].

Decisions regarding therapeutic interventions for hypoglycemic or hyperglycemic episodes are frequently made based on results from these devices, frequently by nurses, pre-hospital emergency medical personnel, and patients[9,10]. Nurses in particular play a fundamental role in the management and care of diabetic patients by relying on the latest guidelines and scientific evidence provided by regulatory bodies. Therefore, we must ensure that nurses understand their key role in monitoring and controlling patients' blood glucose, properly educating patients, and enhancing the necessary skills for patients to measure and interpret their own blood glucose results[11–13]. Nurses must also be aware of potential drawbacks to different testing methodologies and devices. The goal of this paper is to understand whether fingerstick devices are appropriate for meeting the needs of blood glucose monitoring and decision-making.

Background

Nurses assist patients in gaining greater independence in managing their condition through regular and precise self-monitoring of blood glucose, thereby preventing the occurrence of acute and chronic complications of diabetes. This approach ultimately leads to improved treatment adherence, effectiveness of clinical interventions, and, consequently, patient health and quality of life[11,12]. In recent decades, blood glucose self-monitoring has been recognized as a key strategy in diabetes management, as patients who continuously track their blood glucose results are not only more adherent to treatment and self-adjustment of glucose-controlling medications but are also able to take preventive measures when necessary to avoid hypoglycemic events[13]. Existing evidence highlights the significance of this strategy and confirms the undeniable responsibility of nurses in educating, guiding, and supporting patients throughout this process[9,14].

Studies conducted over the past decade on the accuracy and precision of results from these devices have raised questions about the reliability of the data they provide[15]. In a recently published study, glucometers were evaluated in terms of accuracy, precision, level of agreement, and limits of agreement. The findings demonstrated that despite considerable efforts by manufacturers to improve the precision of these devices, their performance still fell below acceptable levels[16]. These errors can lead to incorrect or suboptimal therapeutic decisions, such as inappropriate medication dosage adjustments, which are contributing factors to glycemic emergencies[17]. The evaluation of performance of glucometers is guided by the ISO 15,197 guideline, as recommended by the World Health Organization[18]. However, a large-scale study conducted in 2025 on 56 different brands used by 4,918 participants revealed questionable performance in terms of accuracy and clinical reliability, despite these devices having been certified under ISO 15,197[15]. These findings highlight the necessity for revising and improving the evaluation methods and quality of glucometers in both diabetic and non-diabetic patients[8,16].

Glucometers typically measure blood glucose levels using whole blood samples, whereas laboratory results come from a range of sample types including blood serum, plasma and whole blood[8,19]. Glucometers employ various enzymatic technologies such as glucose oxidase and glucose dehydrogenase for these measurements. This inherent difference in both sample type and enzymatic technology is a major factor contributing to the discrepancies and variability in results between these two methods, as well as among different types of glucometers[20].

This study was designed to evaluate the accuracy, precision, and agreement of a glucometer compared to the standard laboratory method based on ISO 15197:2013 criteria in an Iranian population. The differences in measurement methodologies between different glucometers and between glucometers and standard laboratory tests leaves significant questions as to the trustworthiness of glucometer measurements in guiding treatment decision making. Although relevant guidelines address numerous factors that can influence glucometer measurements—including user technique, hemoglobin levels, medications, and other variables—these factors alone cannot sufficiently account for the systematic differences observed between the two measurement methods[21]. Thus, we argue that there is a need to more rigorously evaluate the accuracy and precision of glucometers in comparison to standard laboratory methods[1,18] to ensure nurses understand the potential challenges these devices create in effective patient care.

Methodology

Study design

This research employed a cross-sectional design and was carried out in June 2025. The reporting of the study was guided by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement[22].

Methods

In this study, 589 participants referred to Dr. Shafaei Pathology Laboratory, both with and without diabetes, were enrolled for blood glucose measurement. Subjects were recruited using a convenience sampling method over a 14-day period during June 2025. Of these, 136 were diabetic and 453 were non-diabetic. Diabetic patients were identified based on self-reported previous medical diagnosis, while non-diabetic participants had no known history of diabetes. The inclusion criteria were as follows: age above 18 years, hemodynamically stable condition, and confirmed diabetic or non-diabetic status. Exclusion criteria included: recent consumption of vitamin C or medications known to cause hyperglycemia, such as corticosteroids, and the presence of any acute stressful condition such as severe infection, recent surgery, or trauma. After assessing eligibility and obtaining informed consent, blood glucose measurement was performed using the venous samples collected for laboratory analysis.

Blood glucose levels were measured using the glucosureSTAR point-of-care glucometer (A. Menarini Diagnostics, Florence, Italy)[23], which employs the glucose oxidase enzymatic technique. Quality control for the glucometer was performed using a normal control solution with a glucose value of 108 mg/dL and an acceptable range of 89–127 mg/dL. Although the device's primary intended use is for capillary whole blood, for the purpose of direct comparison, we used fresh venous whole blood samples collected in clot activator tubes at the time of phlebotomy. This immediate measurement by the glucometer served as the baseline value, minimizing the effect of glycolysis. Within 30–60 min after collection, samples were centrifuged to halt further glucose metabolism by blood cells. This time frame for centrifugation is in accordance with standard laboratory protocols to prevent significant glycolytic loss, which typically occurs at a rate of approximately 5–7 mg/dL per hour (approximately 0.3 to 0.6 mmol/L per hour) in non-centrifuged samples at room temperature. All samples were analyzed in the laboratory within 5 h using a Roche Hitachi 917 analyzer and a glucose oxidase kit (Lot: 17410224) in the laboratory.

Metrological traceability and quality control

Metrological traceability for the reference laboratory method (Roche Hitachi 917) was established through calibration traceable to higher-order reference methods (Isotope Dilution Mass Spectrometry, ID-MS). Internal quality control (IQC) was performed daily using two levels of commercial control materials (Roche Precinorm U and Precipath U) to verify analytical performance. The laboratory also participates in the Iranian External Quality Assessment Scheme (IEQAS). The same primary calibrator batch (Roche Glucose Calibrator, Lot: 17410224) was used for all tests throughout the study to maintain consistency and minimize calibration-related variability.

The glucosureSTAR POCT glucometer (A. Menarini Diagnostics) utilizes factory calibration traceable to a reference method based on YSI (Yellow Springs Instruments) technology. It is recognized that even traceable comparator methods can have different calibration profiles, which is a potential source of systematic difference between the methods.

Demographic and background variables were collected at the same time the phlebotomy sample was collected. These included gender, educational level, marital status, body mass index (BMI), smoking status, alcohol or substance use, and pregnancy status, were recorded. In addition, clinical and contextual variables were documented, such as lipid profile, complete blood count (CBC), hemoglobin level, hematocrit, uric acid level, inflammatory markers, current medications, and ambient temperature and humidity.

Data analysis

Continuous variables were summarized presenting mean and standard deviation, while categorical variables were summarized using frequency and percentage.

All blood glucose measurements—using both the glucometer and the standard laboratory kit—were performed by a single operator. Therefore, the intraclass correlation coefficient (ICC) was calculated to assess the reliability or agreement between the two methodologies. An ICC value above 0.90 is generally considered excellent[24].

Agreement and reliability

The agreement between blood glucose measurements obtained from the glucometer and the standard laboratory kit was evaluated using a Bland-Altman plot. Subsequently, the Coefficient of Repeatability (CR) was calculated, defined as 1.96 times the standard deviation (SD) of the differences between the two measurement methods[25].

Due to the sensitivity of proportional changes between techniques to non-normality, nonlinear relationships, and outliers, agreement was further assessed using Passing-Bablok regression. This is a nonparametric regression method that is robust against outliers and does not assume normality, constant variance ratios, or specific sampling distributions, while accommodating imprecision in both measurement procedures[26]. A prerequisite for applying this method is a strong correlation between measurements, as evaluated by Kendall's Tau coefficient, where an absolute value of $T > |0.7|$ is considered strong[27].

Passing-Bablok regression was used to quantify systematic (intercept A) and proportional (slope B) differences between the two methods. Agreement is indicated if the 95% confidence interval for A contains 0 and for B contains 1.0[28,29]. The cumulative sum (CUSUM) linearity test was applied to assess whether residuals were randomly distributed around the regression line; a p-value ≤ 0.05 suggests significant deviation from linearity, warranting further investigation[26].

Additionally, the concordance correlation coefficient (CCC)[30] was computed to evaluate absolute agreement, assessing the degree to which paired measurements aligned with the line of identity (45° line). The CCC incorporates precision (Pearson's r) and accuracy (bias correction factor Cb). Values were interpreted as follows: >0.99 (almost perfect), 0.95–0.99 (substantial), 0.90–0.95 (moderate), and < 0.90 (poor)[31].

Clinical accuracy assessment

The clinical accuracy of the POCT glucose meter was evaluated by comparing its results against the reference method (laboratory glucose measurement)[32]. The comparison was performed using Parkes (Consensus Error Grid Analysis for Type 1 diabetes)[33] and Clarke Error Grid Analyses (EGA)[34], which are standard methodologies for assessing the clinical significance of glucose measurement errors[32,35].

Analytical accuracy was first assessed against the ISO 15197:2013 criteria [5]. According to these standards, for glucose concentrations < 100 mg/dL, 95% of the POCT results must be within ± 15 mg/dL of the laboratory reference value, and for concentrations ≥ 100 mg/dL, 95% of the results must be within ± 15% of the reference value[18].

Subsequently, clinical accuracy was determined using both the Parkes and Clarke Error Grids. For a device to be clinically acceptable, a minimum of 99% of its data points must fall within clinically acceptable zones (zones A and B) of the respective error grids[18,36].

Precision (repeatability) assessment

The repeatability (intra-assay precision) of the point-of-care glucose meter was assessed in accordance with clause 6.3 of the ISO 15197:2013 standard[33]. A single venous whole blood sample anticoagulated with EthylineDiamine Tetraacetic Acid (EDTA) was used for each test, as recommended for such evaluations[37,38]. Twenty[20] consecutive measurements were performed for each of the three sample pools, which had reference glucose concentrations established by the laboratory method covering low (47 mg/dL), medium (94 mg/dL), and high (165 mg/dL) ranges[18,32]. The entire process for each sample was completed within a short time frame to preclude any potential change in the actual glucose concentration[18,32].

Precision was statistically evaluated by calculating the mean, standard deviation (SD), and coefficient of variation (CV%) for the 20 measurements at each level[18,39]. The Coefficient of Variation (CV%) was calculated for each level using the formula:

CV% = (Standard Deviation/Mean) × 100.

Even though The ISO 15197:2013 standard does not specify particular criteria for measures of precision, yet the acceptance criteria according to conventions as in the study of Ronald Brazg et al.l (2013) for repeatability is stated as a binary pass/fail similar to what follows: a standard deviation of ≤ 5 mg/dL for glucose concentrations less than 100 mg/dL and a coefficient of variation of ≤ 5% for glucose concentrations equal to or greater than 100 mg/dL[40]. Stefan Pleus et al. (2024), set even stricter thresholds for a device to pass the repeatability check[41].

ROC curve analysis for diagnostic accuracy

Receiver operating characteristic (ROC) analysis was performed to evaluate the diagnostic accuracy of the glucometer in distinguishing between diabetic and non-diabetic participants[42]. The analysis yielded measures including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)[43]. The optimal cut-off point for the glucometer readings was determined by maximizing Youden's Index (J), calculated as J = Sensitivity + Specificity – 1[44]. This method identifies the threshold value that best balances the trade-off between true positive and true negative rates, minimizing overall misclassification. The area under the ROC curve (AUC) was computed as a global measure of the glucometer's discriminatory power, where an AUC of 1.0 represents perfect discrimination and 0.5 represents a test no better than chance[45].

All analyses were performed using IBM SPSS Statistics (v26) and MedCalc (v20.015), with statistical significance set at $\alpha = 0.05$ (two-tailed).

Results

Of the 589 participants, there were 406 men (68.9%) and 183 women (31.1%). The mean blood glucose measured by laboratory standard and POCT was 106.40 mg/dL (SD: 38.60) and 109.71 mg/dL (SD: 38.19) for the men and women, respectively. The following sections present analyses separately for participants without and with diabetes.

Method comparison between POCT glucometer and standard laboratory method

The agreement between the POCT glucometer and the standard laboratory method for measuring fasting blood glucose (FBG) was evaluated across three distinct cohorts: (1) individuals without diabetes ($n = 453$), (2) individuals with diabetes on anti-diabetic medication ($n = 136$), and (3) the total combined cohort ($N = 589$). Method comparison was performed using Kendall's Tau correlation, Bland-Altman analysis, Passing-Bablok regression, and CCC (Table 1; Figs. 1 and 2, and 3).

Participants without diabetes

In individuals with no history of diabetes and not using anti-diabetic medications, the results of these analyses are summarized in Table 1; Fig. 1.

Bland-Altman analysis revealed a mean bias of −3.80 mg/dL (95% CI: −5.39 to −2.20), indicating that the POCT glucometer readings were consistently lower than laboratory values. The limits of agreement were wide, ranging from − 35.57 mg/dL (95% CI: −38.30 to −32.84) to 27.97 mg/dL (95% CI: 25.24 to 30.70). Regression analysis of differences versus means showed no significant proportional bias (Slope = −0.08, $p = 0.056$). The coefficient of repeatability was 32.59 (95% CI: 30.48 to 35.03). This indicates that the difference between two

Analysis parameter	Without diabetes	With diabetes	Total cohort
Sample size	453	136	589
Bland-Altman analysis			
Mean bias, mg/dL (95% CI)	-3.80 (-5.39 to -2.20)	-1.64 (-5.90 to 2.60)	-3.30 (-4.87 to -1.74)
Limits of agreement, mg/dL	-35.57 to 27.97	-47.15 to 43.85	-38.66 to 32.04
Proportional bias (Slope, p-value)	-0.08 (p=0.056)	0.028 (p=0.464)	0.01 (p=0.60)
Coefficient of repeatability (95% CI)	32.59 (30.48 to 35.03)	45.42 (40.27 to 52.09)	35.91 (33.85 to 38.25)
Correlation			
Kendall's Tau (95% CI)	0.332 (0.261 to 0.396)	0.668 (0.570 to 0.736)	0.510 (0.435 to 0.541)
Passing-Bablok regression			
Regression equation	POCT = 39.28 + 0.56 × Lab	POCT = 0.17 + 0.98 × Lab	POCT = 18.73 + 0.79 × Lab
Intercept, A (95% CI)	39.28 (32.31 to 45.25)	0.17 (-12.08 to 12.54)	18.73 (12.70 to 24.64)
Slope, B (95% CI)	0.56 (0.50 to 0.63)	0.98 (0.89 to 1.07)	0.79 (0.73 to 0.85)
Residual standard deviation	12.64	16.52	14.29
Cusum test (Linearity)	P = 0.01	P = 0.47	P < 0.01
Concordance analysis			
CCC (95% CI)	0.679 (0.624 to 0.728)	0.916 (0.882 to 0.941)	0.886 (0.866 to 0.903)
Precision (Pearson's ρ)	0.693	0.917	0.890
Accuracy (Cb)	0.980	0.999	0.996

Table 1. Comprehensive method comparison analysis across study cohorts.

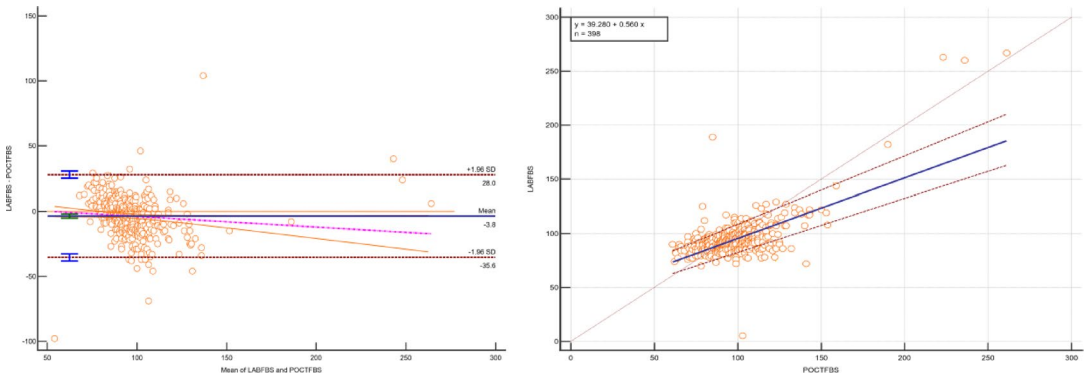

Fig. 1. Bland-Altman plot (left) and Passing-Bablok regression analysis (right) comparing the POCT glucometer and laboratory method in individuals without diabetes and not using anti-diabetic medications ($n = 398$).

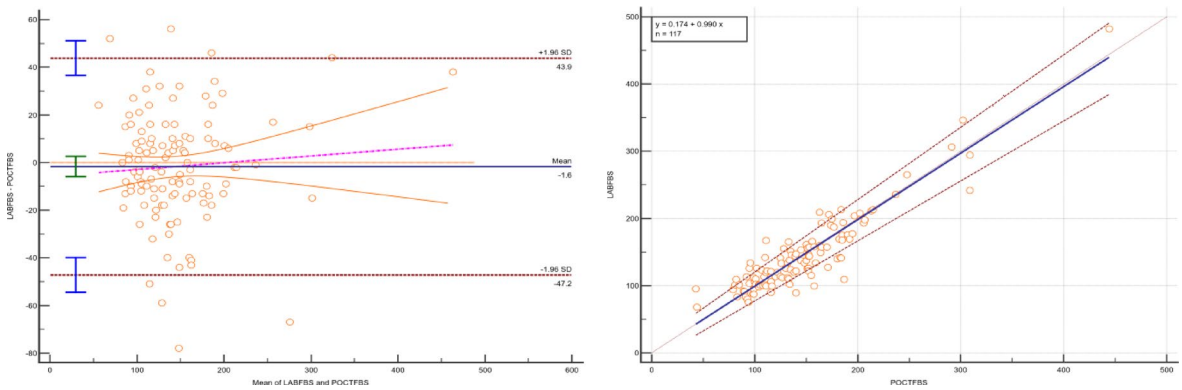

Fig. 2. Bland-Altman plot (left) and Passing-Bablok regression analysis (right) comparing the POCT glucometer and laboratory method in individuals with diabetes using anti-diabetic medications ($n = 117$).

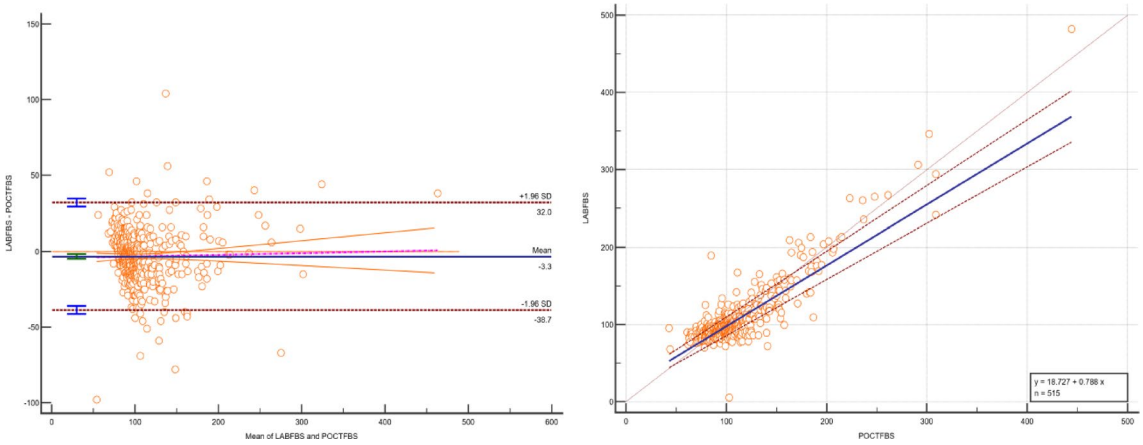

Fig. 3. Bland-Altman plot (left) and Passing-Bablok regression analysis (right) comparing the POCT glucometer and laboratory method in the total cohort of diabetic and non-diabetic participants ($n = 515$).

repeated measurements on the same individual could be as large as 32.59 mg/dL, highlighting substantial variability that is not ideal for a clinical monitoring device.

Kendall's Tau coefficient was 0.332 ($p < 0.0001$; 95% CI: 0.261 to 0.396), indicating the expected statistically significant moderate monotonic relationship between the POCT glucometer and laboratory method.

The Passing-Bablok regression analysis revealed significant systematic differences between the methods, described by the equation: POCT FBS = 39.28 + 0.56 × Lab FBS. The analysis demonstrated a constant error, with an intercept significantly different from zero (A = 39.28; 95% CI: 32.31 to 45.25), indicating a fixed and relatively large bias between the methods. Additionally, a proportional error was observed, as the slope was significantly less than 1 (B = 0.56; 95% CI: 0.50 to 0.63), reflecting an increasing underestimation by the POCT device at higher glucose concentrations. The random differences, quantified by a residual standard deviation of 12.64 (± 1.96 RSD interval: −24.79 to 24.79), indicated considerable variability in the differences between the two methods. Furthermore, the Cusum test for linearity indicated a significant deviation from linearity ($P = 0.01$), suggesting that the relationship between the methods may not be adequately described by a straight line.

The Concordance Correlation Coefficient (CCC) was 0.679 (95% CI: 0.624 to 0.728), indicating poor agreement according to the pre-defined criteria discussed previously. The bias correction factor (Cb = 0.980) confirmed excellent accuracy, while Pearson's ρ (0.693) indicated good precision.

Participants with diabetes

In individuals with diabetes using anti-diabetic medications, the results are summarized in Table 1; Fig. 2.

Bland-Altman analysis showed identical mean bias of −1.64 mg/dL (95% CI: −5.90 to 2.60). The limits of agreement were wide, ranging from − 47.15 mg/dL (95% CI: −54.44 to −39.86) to 43.85 mg/dL (95% CI: 36.57 to 51.14). No significant proportional bias was found (Slope = 0.028, $p = 0.464$). The coefficient of repeatability was 45.42 (95% CI: 40.27 to 52.09). This even larger value signifies exceptionally high measurement variability, where repeated tests could differ by over 45 mg/dL, further underscoring the device's poor reliability for clinical application.

Kendall's Tau coefficient was 0.668 ($p < 0.0001$; 95% CI: 0.570 to 0.736), indicating the expected statistically significant strong monotonic relationship between the methods.

Passing-Bablok regression analysis revealed systematic error patterns between the methods, described by the equation: POCT FBS = 0.17 + 0.98 × Lab FBS. The analysis indicated no constant error, with an intercept of 0.17 with a 95% CI of −12.08 to 12.54, which included zero, suggesting no fixed bias. Additionally, no proportional error was identified, as the slope of 0.98 and 95% CI of 0.89 to 1.07 included 1.0, indicating no systematic underestimation or overestimation across glucose concentrations. The random differences, quantified by a residual standard deviation of 16.52 (± 1.96 RSD interval: −32.38 to 32.38), indicated considerable variability between the two methods. Furthermore, the Cusum test for linearity showed no significant deviation from linearity ($P = 0.47$), confirming the appropriateness of the linear model for describing the relationship between the methods.

The Concordance Correlation Coefficient (CCC) was 0.916 (95% CI: 0.882 to 0.941) indicating stronger agreement than in the non-diabetes group but poor agreement overall. There was excellent accuracy (Cb = 0.999) and good precision (Pearson's ρ = 0.917).

Total data

In the combined cohort of diabetic and non-diabetic participants (Table 1; Fig. 3):

Bland-Altman analysis

Bland-Altman analysis showed identical mean bias of −3.30 mg/dL (95% CI: −4.87 to −1.74). The limits of agreement were wide, ranging from − 38.66 mg/dL (95% CI: −41.34 to −35.99) to 32.04 mg/dL (95% CI: 29.37

to 34.71). No significant proportional bias was found (Slope = 0.01, $p = 0.60$). The coefficient of repeatability was 35.91 (95% CI: 33.85 to 38.25).

Kendall's Tau correlation

Kendall's Tau coefficient was 0.510 ($P < 0.0001$; 95% CI: 0.435 to 0.541), indicating a statistically significant moderate monotonic relationship.

Passing-Bablok regression

Passing-Bablok regression analysis yielded the equation POCT FBS = 18.73 + 0.79 × Lab FBS, revealing significant systematic discrepancies between the methods. A constant error was evident from an intercept of 18.73 (95% CI: 12.70 to 24.64), significantly differing from zero and indicating a fixed bias. Additionally, a proportional error was identified, with a slope of 0.79 (95% CI: 0.73 to 0.85) significantly less than 1, reflecting an increasing underestimation by the POCT device at higher glucose concentrations. Random differences were substantial, as indicated by a residual standard deviation of 14.2868 (± 1.96 RSD interval: −28.0021 to 28.0021). Furthermore, the Cusum test demonstrated a significant deviation from linearity ($P < 0.01$), suggesting that the relationship between the methods may not be adequately described by a linear model.

Concordance correlation coefficient

The CCC was 0.886 (95% CI: 0.866 to 0.903), classified as poor agreement. The bias correction factor (Cb = 0.996) indicated excellent accuracy, while Pearson's ρ (0.890) showed good precision.

Clinical accuracy results

A total of 589 paired glucose measurements were obtained and analyzed for clinical and analytical accuracy. The analytical performance of the point-of-care glucose meter was assessed against the ISO 15197:2013 standards. The results demonstrated that 95.1% (560 out of 589) of the measurements met the stipulated criteria (± 15 mg/dL for values < 100 mg/dL and ± 15% for values ≥ 100 mg/dL), thereby fulfilling the minimum ISO requirement of 95%.

The clinical significance of the measurement differences was evaluated using error grid analysis.

- In the Parkes Error Grid (for Type 1 diabetes), 100% of the data points (589 out of 589) resided in clinically acceptable zones (Zones A and B). The distribution was 97.7% in Zone A (no effect on clinical action) and 2.3% in Zone B (altered clinical action with little or no effect on outcome). No points were found in Zones C, D, or E.
- Analysis with the Clarke Error Grid showed that 99.8% of results (588 out of 589) were in zones A and B. One data point (0.2%) was located in Zone D, which would indicate a dangerous failure to detect hypoglycemia or hyperglycemia (Table 1).

Precision (repeatability) assessment

The repeatability (precision) of the glucose meter was evaluated across three different glucose concentrations. At the low glucose level (reference value: 47 mg/dL), the mean POCT measurement was 38.2 mg/dL with a SD of 4.52 mg/dL and a CV% of 11.83%. The SD met the ISO 15197:2013 requirement for concentrations below 100 mg/dL (SD ≤ 5 mg/dL). At the medium glucose level (reference value: 94 mg/dL), the mean POCT value was 93.1 mg/dL with an SD of 10.73 mg/dL and a CV% of 11.53%. The SD at this level exceeded the ISO limit of ≤ 5 mg/dL for concentrations < 100 mg/dL, indicating a failure to meet precision requirements. At the high glucose level (reference value: 165 mg/dL), the mean POCT measurement was 177.3 mg/dL with an SD of 9.43 mg/dL and a CV% of 5.32%. Although the SD was not the primary criterion at this concentration, the CV% of 5.32% marginally exceeded the ISO requirement of ≤ 5% for values ≥ 100 mg/dL. These results indicate that the precision of the glucose meter meets the ISO 15197:2013 requirement at the low concentration level but fails to meet the required precision criteria at the medium and high concentration levels under the tested conditions (Table 2).

ROC curve analysis for optimal cut-off determination

The diagnostic performance of the blood glucose test was assessed using ROC curve analysis to determine its efficacy in discriminating between diabetic and non-diabetic individuals (Fig. 4). For this specific analysis, the cohort was refined to create distinct diagnostic groups. Participants with prediabetes ($n = 51$) were excluded to

Glucose level	Reference value (mg/dL)	Mean (SD) POCT (mg/dL)	Coefficient of variation (CV%)	ISO 15197:2013 CV requirement*
Low	47	38.2 (4.52)	11.83%	≤ SD: 5 mg/dL for x < 50 mg/dL
Medium	94	93.1(10.73)	11.53%	≤ CV: 5% for x < 100 mg/dL
High	165	177.3 (9.43)	5.32%	≤ CV: 5% for x ≥ 100 mg/dL

Table 2. Repeatability analysis of the glucose meter across three glucose concentrations (n=20 measurements per level). *Note: The ISO 15197:2013 standard specifies that for repeatability, the SD should be ≤ 5 mg/dL for concentrations < 100 mg/dL and CV ≤ 5% for concentrations ≥ 100 mg/dL. The medium level (94 mg/dL) is below the 100 mg/dL threshold for the CV requirement, so the SD criterion is more appropriate. Its SD (10.73) is compared to the 5 mg/dL limit

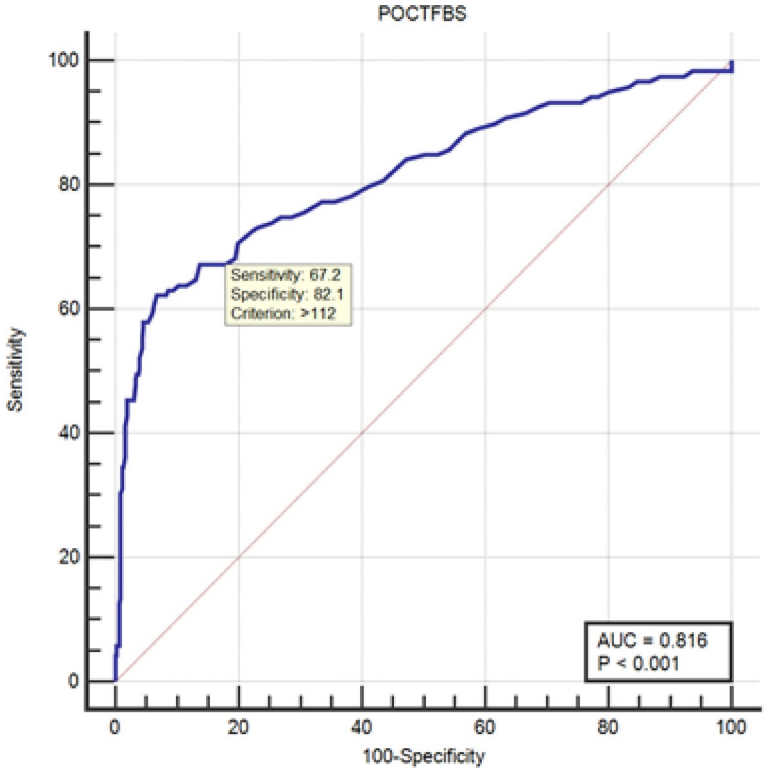

Fig. 4. Receiver operating characteristic (ROC) curve analysis of the glucosureSTAR glucometer for discriminating between diabetic and non-diabetic participants.

ensure a clear comparison between confirmed diabetic and confirmed non-diabetic individuals. Consequently, the analysis was conducted on a total cohort of 549 subjects, comprising 119 (21.68%) diabetic patients and 430 (78.32%) non-diabetic participants.

The overall accuracy of the test, as quantified by the AUC, was 0.816 (95% CI: 0.781 to 0.847; Standard Error: 0.025). This AUC was found to be statistically significant ($p < 0.0001$), confirming that the test possesses a good level of discriminatory power significantly better than chance (AUC = 0.5).

To identify the most suitable cut-off value for clinical use, the Youden Index (J) was calculated to maximize the sum of sensitivity and specificity. The maximum Youden Index was 0.5544, which corresponded to a blood glucose criterion of > 124 mg/dL. This value was therefore selected as the optimal diagnostic threshold.

At that cut point, the test demonstrated a sensitivity of 62.18% (95% CI: 52.8 to 70.9) and a specificity of 93.26% (95% CI: 90.5 to 95.4). The likelihood ratios, which express the odds of a given test result occurring in a patient with the condition versus without it, were also calculated. The positive likelihood ratio (+LR) at this threshold was 9.22 (95% CI: 6.32 to 13.46), and the negative likelihood ratio (-LR) was 0.41 (95% CI: 0.32 to 0.51).

Discussion

The goal of this study was to evaluate the technical performance of the glucosureSTAR POCT glucometer as compared to a "gold standard" laboratory test, so that nurses and other health care providers can be aware of their potential shortcomings. The results reveal a mixed performance profile characterized by acceptable clinical risk assessment but suboptimal analytical accuracy and limited diagnostic utility. Our findings contribute to the growing body of literature expressing concern over the reliability of many commercially available glucose meters, even those conforming to broadly accepted international standards[21]. The observed discrepancy between marginal analytical performance and acceptable clinical risk categorization underscores the critical need for more stringent evaluation protocols, and clear communication to the health care providers and patients using these devices.

The fundamental finding of this study is that while 95.1% of results technically met the ISO 15197:2013 analytical criteria, this marginal pass rate—exceeding the minimum requirement by only 0.1%—masks notable measurement variability and bias that can have serious clinical implications. Several findings raise questions about the accuracy and reliability of measurements from the glucometer. First, a negative bias of −3.3 mg/dL ($p < 0.0001$) indicates systematic underestimation of glucose levels. Second, the broad limits of agreement (−38.7 to 32.1 mg/dL), demonstrate substantial measurement scatter (unreliability). Third, the poor concordance correlation coefficient (CCC = 0.886), indicates inadequate absolute agreement with the reference lab method[30,46].

The clinical significance of this imprecision becomes apparent when considering insulin dosing calculations. For a patient with a true glucose value of 180 mg/dL, the observed limits of agreement could yield meter readings

ranging from approximately 141 to 212 mg/dL—a range spanning from near-normal to hyperglycemic glucose concentrations that could lead to dangerously inappropriate therapeutic decisions[36].

On a broader policy level, our finding that the device marginally passed ISO accuracy criteria (95.1%) highlights a fundamental weakness in a binary pass/fail system that does not penalize for systematic bias or wide variability. A device that consistently underestimates glucose levels by a fixed margin could still technically pass the ISO standard, yet pose a significant clinical risk by consistently guiding therapy towards hypoglycemia. The broad range of scores show that repeated measurements could yield markedly different results from the same individual with the same glucose levels, raising questions about whether treatments will match the actual blood glucose levels. This suggests that regulatory standards should be augmented with metrics that quantify systematic error and its potential clinical impact, moving beyond simple percentage-within-limits assessments[19].

Discrepancies between diabetic and non-diabetic patients

The results above suggest some discrepancies in performance as a function of whether the individual being measured is classified as diabetic or not. The superior agreement (CCC = 0.916) and absence of significant systematic bias in diabetic patients suggest that the device's algorithm may be optimized for the physiological milieu of diabetes, such as altered hematocrit levels or increased erythrocyte turnover, which are common in this population. However, these results raise questions about use in nondiabetic populations. In this nondiabetic population, there was substantial bias, a relatively high coefficient of repeatability, and low CCC, suggesting that these devices may be unsuitable for this population. More broadly, these findings highlight how variable device performance can be across patients with different physiologies. Future research should examine whether other physiological differences (ethnicity, age, sex, etc.) could also cause significant variability in performance, which is a critical consideration for clinical use[20].

The paradox of clinical acceptability despite analytical deficiencies

The paradoxical finding of 100% clinical acceptability on Parkes error grid analysis despite analytical deficiencies can be explained by the error grid's design, which prioritizes clinical outcome over analytical precision[33]. This grid categorizes readings based on their potential to cause clinical harm rather than their numerical accuracy. However, the presence of a single reading (0.2%) in zone D of the Clarke error grid—representing a potentially dangerous failure to detect hypoglycemia or hyperglycemia—serves as a crucial reminder that even minimal error rates can have severe consequences in clinical practice[34,47]. This finding suggests that while error grid analysis remains valuable for assessing clinical risk, it should complement rather than replace rigorous analytical validation.

The single reading in Clarke error grid Zone D, while a small percentage, is clinically indefensible. In a real-world setting, this single error could represent a catastrophic failure to detect severe hypoglycemia, potentially leading to a fatal outcome. This underscores the argument that for devices used in making critical therapeutic decisions, a 'six sigma' approach with near-zero error rates for dangerous misclassifications should be the benchmark, rather than the current ISO minimums[47].

It is essential that independent evaluation establishes the precision, accuracy, and reliability of devices like these glucometers in "real world" usage with all intended subgroup populations, provided their screening utility in low-resource areas, guiding our decision making particularly in prehospital emergencies or providing proper basis for further referrals to make in home-care and SMBG. Failure to provide accurate, reliable instruments for identifying and/or managing diseases like diabetes can pose significant cumulative costs of not managing this complicated metabolic disease both for individuals and at the population level. The device's particularly concerning diagnostic performance, with a sensitivity of only 62.18% at the optimal cut-off, falls substantially short of established benchmarks for screening tests. Authoritative sources, such as the Clinical and Laboratory Standards Institute (CLSI) guideline EP12-A2, suggest that for a test to be considered adequate for screening purposes—where the cost of missing true cases is high—sensitivity should ideally be 90% or greater[48,49]. The observed sensitivity of 62.18% would fail to identify approximately 38% of individuals with dysglycemia, creating a dangerous potential for delayed diagnosis and management[49]. This low sensitivity renders it inadequate for screening purposes, as it would miss approximately 38% of true cases of dysglycemia. The high specificity (93.26%), while favorable for confirming positive results, does not compensate for this critical deficiency in ruling out disease. The area under the ROC curve (AUC = 0.816), while statistically better than chance, falls short of the excellence (AUC > 0.90) required for a reliable diagnostic tool in clinical practice[42,50].

The positive likelihood ratio (+LR) of 9.22 signifies that a glucose reading above the 124 mg/dL cut-off is approximately nine times more likely to be observed in a diabetic individual than in a non-diabetic one. This represents a moderately large, and often clinically important, shift in pre-test probability toward ruling in dysglycemia. Conversely, the negative likelihood ratio (-LR) of 0.41 indicates that a negative test result is only about 0.4 times as likely in a person with diabetes, resulting in only a small to moderate decrease in the probability of disease. Therefore, while the strong +LR is a notable strength for confirming dysglycemia, the less impressive -LR greater than 0.1 underscores the device's limitation in reliably ruling out the condition, a finding that is consistent with its suboptimal sensitivity.

Several technical factors may explain the observed performance limitations. The systematic underestimation could originate from the sample matrix difference (whole blood in glucometers versus plasma in laboratory analyzers), which typically requires a conversion factor of approximately 1.11–1.12 that may not be optimally calibrated in this device[38]. Although the glucosureSTAR manufacturer claims the device includes hematocrit compensation, the significant variability observed, particularly across our diverse patient population, suggests potential under-compensation or an imperfect calibration of this feature. Hematocrit variability remains a well-documented interferent in glucose oxidase-based systems and likely contributed to the random variability observed in this study[20].

Additionally, hematocrit variability—a well-documented interferent in glucose oxidase-based systems—likely contributed to the random variability observed, particularly given the diverse patient population included in this study[20]. The device's failure to meet precision requirements at medium and high glucose concentrations further suggests limitations in the enzymatic reaction kinetics or electrochemical detection system at higher analyte concentrations[19].

It should be noted that the glucometer was used with venous whole blood instead of its primarily intended capillary sample. This methodological choice, while necessary for simultaneous comparison with the laboratory analyzer, may have influenced the observed measurement bias and variability. An important limitation of this study is the use of venous whole blood samples for the glucometer analysis. According to the manufacturer's Instructions for use, the glucosureSTAR device is calibrated for use with capillary whole blood. The use of venous blood, which may differ in hematocrit and oxygen content, could be a contributing factor to the systematic bias and wide limits of agreement observed in our results. Future studies should utilize the intended capillary sampling method to validate these findings.

A further consideration is the potential for calibration profile differences between the two systems. The glucometer's calibration is traceable to a YSI-based method, while the laboratory analyzer is traceable to ID-MS. As noted in the literature, such differences in the calibration of comparator methods, despite both being traceable to higher-order standards, can contribute to systematic bias.

The device's failure at higher glucose concentrations is particularly concerning for its intended use in diabetic populations, where hyperglycemia is a primary target of management. This failure likely reflects a limitation in the enzyme-electrode chemistry, such as substrate saturation or interference from alternative electron acceptors at high glucose levels. This technological limitation directly impacts its utility for titrating therapy in poorly controlled diabetics, who need accurate measurements precisely in this high range[21].

The implications of these findings extend beyond this specific device to highlight broader concerns in POCT glucose monitoring. Regulatory approval based on meeting minimum ISO standards does not necessarily guarantee clinical reliability across all use scenarios[50]. The contrast between acceptable error grid performance and marginal analytical accuracy suggests that current regulatory frameworks may prioritize the avoidance of extreme errors over ensuring consistent precision across the measuring range.

Our study has several design limitations that should be considered when interpreting these results. The use of a convenience sampling method, while practical, may have introduced selection bias. Although we controlled for numerous confounding variables including hematocrit, lipid levels, and medications, unmeasured factors such as altitude variations, abnormal oxygen tension, or rare interfering substances could have influenced the results. The single-operator design, while strengthening internal consistency for method comparison, may limit generalizability to real-world settings with multiple users of varying expertise. However, given this glucometer is a medical device intended for use in the general population, the results are not easily explained by potential sampling bias. Furthermore, the single-operator design should have served to limit possible error variance associated with multiple operators in multiple locations potentially testing in different environments. Thus, these results may actually underestimate the severity of the issue when real-world usage is considered.

In conclusion, the glucosureSTAR glucometer demonstrates a performance profile that necessitates cautious and context-specific application. Its acceptable clinical risk profile may permit its use for trend monitoring in clinically stable patients with established diabetes who understand its consistent negative bias and variability. However, its marginal analytical performance, significant systematic error, and poor diagnostic sensitivity categorically preclude its recommendation for screening, diagnosis, or making fine adjustments to therapy. These findings emphasize the critical responsibility of regulatory bodies, manufacturers, and clinical professionals to demand more rigorous analytical performance and real-world diagnostic validation beyond minimum ISO standards and error grid analyses. Future developments should focus on improving hematocrit independence, reducing matrix-related biases, and enhancing precision across the entire measuring range to ensure patient safety and effective diabetes management in diverse clinical scenarios[50].

Relevance to clinical practice

This study highlights critical limitations in the performance of the glucosureSTAR glucometer, underscoring serious concerns for diagnostic or screening purposes due to low sensitivity, systematic bias, and significant analytical variability. However, with appropriate user awareness of its consistent underestimation bias and wide limits of agreement, it may be cautiously used for trend monitoring in known diabetic patients in settings where rapid results are prioritized over absolute accuracy, such as emergency or field conditions. Healthcare providers, particularly nurses and emergency personnel, should be trained to interpret results within the device's limitations and avoid relying on it for therapy adjustment or initial diagnosis. These findings advocate for stricter device validation and context-specific guidelines to ensure patient safety in point-of-care glucose monitoring.

Conclusion

This comprehensive evaluation reveals that the glucosureSTAR POCT glucometer demonstrates a performance profile requiring careful contextual application. While the device showed acceptable clinical risk categorization with 100% of Parkes error grid points in zones A/B, several analytical deficiencies limit its utility. The consistent underestimation bias (−3.3 mg/dL), wide limits of agreement (−38.7 to 32.1 mg/dL), poor concordance (CCC = 0.886), and failure to meet precision standards at medium and high glucose concentrations indicate substantial measurement variability. Furthermore, the suboptimal diagnostic performance (AUC = 0.816; sensitivity = 62.18% at > 124 mg/dL cut-off) renders it unreliable for screening or diagnosis.

These findings suggest that while the glucometer may be cautiously employed for trend monitoring in clinically stable patients with established diabetes, particularly in emergency settings where rapid results

outweigh precision and accuracy needs, it is not recommended for diagnosis, therapy titration, or any scenario requiring high analytical accuracy.

The marginal pass rate of ISO standards (95.1%) given the problematic results highlights the critical need for more rigorous device validation beyond minimum regulatory requirements. Healthcare providers should be adequately trained to interpret results within the device's limitations and understand its systematic underestimation bias. This study underscores the necessity for continuous quality improvement in POCT devices and context-specific validation to ensure patient safety and effective diabetes management across diverse clinical scenarios.

Data availability

The data supporting the findings of this study can be obtained from the corresponding author upon request. Please feel free to reach out for access to the data.

Received: 8 September 2025; Accepted: 7 November 2025
Published online: 25 November 2025

References

1. Weinstock, R. S. et al. The role of blood glucose monitoring in diabetes management. (2020).
2. Gohumpu, J., Lim, W. K., Peng, Y., Xue, M. & Hu, Y. (eds) Enhancing user experience: innovations in blood glucose meter design for improved efficiency and convenience. International Conference on Human-Computer Interaction; : Springer. (2024).
3. Hina, A. & Saadeh, W. Noninvasive blood glucose monitoring systems using near-infrared technology-a review. *Sens. (Basel)* **22**(13), 4855 (2022).
4. Mata-Cases, M. et al. The association between poor glycemic control and health care costs in people with diabetes: A Population-Based study. *Diabetes Care*. **43** (4), 751–758 (2020).
5. D'Antoni, F. et al. Layered meta-learning algorithm for predicting adverse events in type 1 diabetes. *IEEE Access*. **11**, 9074–9094 (2023).
6. Goldstein, D. E. et al. Tests of glycemia in diabetes. *Diabetes Care*. **27** (7), 1761–1773 (2004).
7. Woldaregay, A. Z. et al. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. *J. Med. Internet. Res.* **21** (5), e11030 (2019).
8. Olansky, L. & Kennedy, L. Finger-stick glucose monitoring: issues of accuracy and specificity. *Diabetes Care*. **33** (4), 948–949 (2010).
9. Mathew, T. K. & Tadi, P. Blood glucose monitoring. (2020).
10. Schwerin, D. L. & Svancarek, B. EMS Diabetic Protocols For Treat and Release: StatPearls Publishing, Treasure Island (FL); 2025. (2025).
11. Awang Ahmad, N. A., Sallehuddin, M. A. A., Teo, Y. C. & Abdul Rahman, H. Self-Care management of patients with diabetes: nurses' perspectives. *J. Diabetes Metabolic Disorders*. **19** (2), 1537–1542 (2020).
12. Carrera, C. M. T. et al. Diabetes nursing education its implication towards an improved quality of life of persons with diabetes: A systematic review. *World* **3**, 1142 (2024).
13. Renard, E. Monitoring glycemic control: the importance of self-monitoring of blood glucose. *Am. J. Med.* **118** (9), 12–19 (2005).
14. Dailah, H. G. (ed) Editor the influence of nurse-led interventions on diseases management in patients with diabetes mellitus: a narrative review (MDPI, 2024).
15. Ukpe, M. P. & Ezeanuka, A. C. Assessment of accuracy, clinical validity, and analytical linearity in point-of-care glucose monitoring devices for diabetes mellitus: A systematic review and meta-analysis. *Clin. Biochem.* **137**, 110911 (2025).
16. Gümüş, A. et al. Glucometer versus analyzer: comparable results with negligible clinical risk. *Scand. J. Clin. Lab. Investig.* **85**(1), 11–9 (2025).
17. Pardo, S. & Parkes, J. L. Predicted blood glucose from insulin administration based on values from miscoded glucose meters. *J. Diabetes Sci. Technol.* **2**(4), 557–562 (2008).
18. Jendrike, N., Baumstark, A., Kamecke, U., Haug, C. & Freckmann, G. ISO 15197: 2013 evaluation of a blood glucose monitoring system's measurement accuracy. *J. Diabetes Sci. Technol.* **11** (6), 1275–1276 (2017).
19. Baumstark, A. et al. Lot-to-Lot variability of test strips and accuracy assessment of systems for Self-Monitoring of blood glucose according to ISO 15197. *J. Diabetes Sci. Technol.* **6** (5), 1076–1086 (2012).
20. Tonyushkina, K. & Nichols, J. H. Glucose meters: a review of technical challenges to obtaining accurate results. *J. Diabetes Sci. Technol.* **3** (4), 971–980 (2009).
21. 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.* **13** (3), 575–583 (2019).
22. Lachat, C. et al. Strengthening the reporting of observational studies in Epidemiology–nutritional epidemiology (STROBE-nut): an extension of the STROBE statement. *Nutr. Bull.* **41** (3), 240–251 (2016).
23. Diagnostics, M. Menarini Diagnostics Official Website - Search Page 2024 [Available from: https://www.menarinidiagnostics.com/en/search.html
24. Sharif Nia, H. et al. The factor structure of the spiritual well-being scale in veterans experienced chemical weapon exposure. *J. Relig. Health*. **57** (2), 596–608 (2018).
25. Bland, J. M. & Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327** (8476), 307–310 (1986).
26. Bilic-Zulle, L. Comparison of methods: passing and Bablok regression. *Biochemia Medica: Biochemia Med.* **21** (1), 49–52 (2011).
27. Akoglu, H. User's guide to correlation coefficients. *Turkish J. Emerg. Med.* **18** (3), 91–93 (2018).
28. Passing, H. & Bablok, W. Comparison of several regression procedures for method comparison studies and determination of sample sizes application of linear regression procedures for method comparison studies in clinical chemistry, part II. (1984).
29. Sharif Nia, H. et al. Clinical accuracy and agreement between tympanic and forehead body temperature measurements for screening of patients with COVID-19. *J. Clin. Nurs.* **31** (21–22), 3272–3285 (2022).
30. Lawrence, I. & Lin, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* https://doi.org/10.2307/2532051 (1989).
31. McBride, G. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA client report: HAM2005-062. ;62. (2005).
32. Guidance, D. Blood glucose monitoring test systems for prescription point-of care use: draft guidance for industry and food and drug administration staff. *Fed. Regist.* **1988**, 1–38 (2005).
33. Parkes, J. L., Slatin, S. L., Pardo, S. & Ginsberg, B. H. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*. **23** (8), 1143–1148 (2000).

34. Clarke, W. L., Cox, D., Gonder-Frederick, L. A., Carter, W. & Pohl, S. L. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*. **10** (5), 622–628 (1987).
35. Pfützner, A., Klonoff, D. C., Pardo, S. & Parkes, J. L. Technical aspects of the Parkes error grid. *J. Diabetes Sci. Technol.* **7** (5), 1275–1281 (2013).
36. Boyd, J. C. & Bruns, D. E. Quality specifications for glucose meters: assessment by simulation modeling of errors in insulin dose. *Clin. Chem.* **47** (2), 209–214 (2001).
37. Abdulla, H. et al. Physiological mechanisms of action of incretin and insulin in regulating skeletal muscle metabolism. *Curr. Diabetes. Rev.* **10** (5), 327–335 (2014).
38. D'Orazio, P. et al. Approved IFCC recommendation on reporting results for blood glucose: international federation of clinical chemistry and laboratory medicine scientific Division, working group on selective electrodes and point-of-care testing (IFCC-SD-WG-SEPOCT). *Clin. Chem. Lab. Med. (CCLM)*. **44** (12), 1486–1490 (2006).
39. Wayne, P. Evaluation of precision of quantitative measurement procedures; Approved guideline. CLSI document EP05-A3[Google Scholar]. (2014).
40. Brazg, R. et al. Clinical evaluation of the freestyle precision pro system. *Clin. Chim. Acta*. **421**, 243–250 (2013).
41. Pleus, S. et al. Evaluation of system Accuracy, Precision, hematocrit Influence, and user performance of two blood glucose monitoring systems based on ISO 15197: 2013/EN ISO 15197: 2015. *Diabetes Therapy*. **15** (2), 447–459 (2024).
42. Zhou, X-H., Obuchowski, N. A. & McClish, D. K. Statistical methods in diagnostic medicine (Wiley, 2011).
43. Altman, D. G. & Bland, J. M. Statistics notes: diagnostic tests 2: predictive values. *Bmj* **309** (6947), 102 (1994).
44. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3** (1), 32–35 (1950).
45. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143** (1), 29–36 (1982).
46. McBride, G. A proposal for strength-of-agreement criteria for lin's concordance correlation coefficient. *NIWA Client Report: HAM2005-062*. **45**, 307–310 (2005).
47. Stockl, D., Dewitte, K., Fierens, C. & Thienpont, L. M. Evaluating clinical accuracy of systems for Self-Monitoring of blood glucose by error grid analysis. Diabetes Care ;**23**(11). (2000).
48. Clinical, Institute, L. S. Evaluation of qualitative, binary output examination performance (Clinical and Laboratory Standards Institute Berwyn, PA, 2023).
49. Sacks, D. B. et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Clin. Chem.* **57** (6), e1–e47 (2011).
50. Klonoff, D. C. et al. Investigation of the accuracy of 18 marketed blood glucose monitors. *Diabetes Care*. **41** (8), 1681–1688 (2018).

Acknowledgements

We express our sincere gratitude to all the patients who took part in this study and generously shared their valuable experiences with us. Additionally, we extend our gratitude to the members of the Dr Shahryar Shafaei pathology laboratory center for their generous assistance and cooperation in recruiting eligible participants and gathering pertinent data. We utilized Artificial Intelligence tools (ChatGPT-4,5 and DeepSeek-V3.1) solely to provide support for improving grammar and readability, not for generating the content of this article, without violating the rights of confidentiality. This assistance resides in the limitations imposed by the journal. The authors verified the outputs and take full responsibility for the final product.

Author contributions

Performance of data gathering: MS; Planning and supervision of the work: HSH; Performance of the analysis: HSH and JO, Manuscript draft: MS, HSH and All authors; and comment on the final manuscript: JO and All authors.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

This study was conducted in accordance with the Declaration of Helsinki and received approval from the Ethics Committee of Mazandaran University of Medical Sciences in Sari, Iran (code: IR.MAZUMS.REC.1404.133). Prior to data collection, participants received detailed information about the study's objectives, procedures, potential risks, and benefits. They were given the opportunity to ask questions before providing written informed consent. For individuals unable to consent independently, written consent was obtained from their legal guardians. The consent forms emphasized voluntary participation, confidentiality, anonymity, and the right to withdraw at any time without consequences. All procedures adhered to relevant ethical guidelines, including those set by the World Medical Association. Permissions to use data collection tools were obtained from their developers. The study was designed to minimize risks while safeguarding participant rights and well-being. Data were handled in compliance with institutional guidelines, ensuring secure storage and transparency regarding data usage in research dissemination.

Additional information

Correspondence and requests for materials should be addressed to M.S.