Glucose Testing Methods: A Systematic Review and Meta-Analysis of Diagnostic Accuracy of Point-of-Care Devices for Neonatal Hypoglycemia

Sophie L. St Clair[1], Caitlyn M. Ulyatt[1], Maria T. Corkin, PhD[1,2], Libby G. Lord[1], Caroline A. Crowther, MD[1], Jane E. Harding, DPhil[1], and Luling Lin, PhD[1]

Objective To evaluate the accuracy of various point-of-care device methodologies for measuring blood glucose concentrations in babies at risk of neonatal hypoglycemia.

Study design This systematic review and meta-analysis included studies from Ovid MEDLINE, Ovid Embase, and Web of Science up to May 20, 2024. Studies comparing point-of-care testing methods for neonatal blood glucose to a standard laboratory method were included, excluding those on continuous glucose monitoring or conducted before 1990. Two researchers independently assessed inclusion and evaluated risk of bias using QUADAS-2. Sensitivity and specificity were calculated using contingency tables, and diagnostic accuracy was analyzed using hierarchical random-effects modelling. Studies with insufficient data were summarized by estimation direction.

Results Seventy-one studies were included. The quantitative analysis (n = 31) evaluated glucose oxidase (GO) + photometry (n = 8), glucose-1-dehydrogenase (GDH) + photometry (n = 6), GO + electrochemistry (n = 13), GDH + electrochemistry (n = 12), and hexokinase (HK) + electrochemistry (n = 2). All methods showed high specificity (≥93%), with GO + electrochemistry, GDH + electrochemistry, and HK + electrochemistry showing superior sensitivity. The summary receiver operating characteristic curve confirmed HK + electrochemistry as the most accurate.

Conclusion Certain point-of-care device methodologies demonstrate greater accuracy in measuring neonatal blood glucose concentrations. Of the methods evaluated, HK + electrochemistry proved to be the most reliable. However, the limited number of studies using this method suggests the need for further research to confirm these findings across diverse settings and populations. *(J Pediatr 2025;278:114438)*.

Systematic review registration PROSPERO on December 29, 2023 (CRD42023488539).

Hypoglycemia is a common neonatal condition that affects up to 15% of all infants[1] and half of "at-risk" infants,[2] which includes those born late-preterm (35-36 weeks' gestation), small or large birth weight, or having a mother with diabetes.[3] The known association of neonatal hypoglycemia with neurodevelopmental impairment[4,5] warrants prompt treatment of any detected episodes. However, hypoglycemia is commonly asymptomatic, so it is recommended that at-risk infants undergo screening by regular measurement of blood glucose concentrations in the first days after birth.[6]

Blood glucose concentrations can be measured by testing the infant's arterial, venous, or capillary blood sample, using a point-of-care method, or using standard laboratory methods (usually glucose oxidase or hexokinase methods). Although laboratory methods are the diagnostic standard and have a high degree of accuracy,[7,8] the turnaround time of results potentially delays the timely treatment of low blood glucose concentrations.[9] By comparison, point-of-care testing methods can be conducted near to where the infant is being treated, allowing for rapid results and prompt management decisions.[9] However, concerns have been raised about the inaccuracies of point-of-care devices, which may result in cases of neonatal hypoglycemia going undetected, or unnecessary treatment of infants with normal blood glucose concentrations.[10] For instance, environment (humidity, temperature)[11] degradation of supplies, preparation and handling of samples,[12] and hematocrit levels, which are highly variable in neonates[13] can affect results. Further, point-of-care devices have been developed primarily for measurement of high blood glucose concentrations in adults with diabetes, and tend to be less accurate at the lower blood glucose concentrations of relevance in the neonatal population.[14] A lack of consensus also remains around whether the results of point-of-care testing methods alone can be used to inform management of neonatal hypoglycemia, and even whether point-of-care methods are reliable enough to serve as screening tests to determine which samples require laboratory confirmation.[7]

HK	Hexokinase
GO	Glucose oxidase
GDH	Glucose-1-dehydrogenase
SROC	Summary receiver operating characteristic

From the [1]Liggins Institute, The University of Auckland, Auckland, New Zealand; and [2]Faculty of Science, School of Psychology, The University of Auckland, Auckland, New Zealand

https://doi.org/10.1016/j.jpeds.2024.114438

A wide range of point-of-care devices are now available for measuring blood glucose concentrations, employing a range of testing methods, including electrochemical[7] or photometric measurement[15] and reaction enzymes.[16] Point-of-care devices that employ enzymatic methods may vary in which reaction enzyme is used (eg, glucose oxidase, glucose dehydrogenase, or hexokinase), and some devices use a combination of enzymatic and photometric or electrochemical methods. However, little guidance exists on which methods may be most appropriate for neonates.[17]

To help address this, this review aimed to determine the accuracy of point-of-care devices for detecting hypoglycemia in neonates.

Methods

The study protocol for this review was registered on PROSPERO on 29/12/2023 (CRD42023488539). The study was conducted according to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[18] and reported according to the PRISMA extension for Diagnostic Test Accuracy studies[19] (**Appendix 1**; available at www.jpeds.com).

Search Strategy and Study Selection

We searched Ovid MEDLINE, Ovid Embase, and Web of Science, without language restriction, from inception to 20 May 2024 (**Appendix 2**; available at www.jpeds.com). At the full-text review stage, publication dates were restricted to 1990 and onwards to ensure only currently relevant testing methods were included. Conference abstracts were included if they provided sufficient data but were merged with their corresponding full text if this was available. There were no restrictions on study type.

Eligible studies compared a point-of-care method for testing blood glucose concentration (utilizing any blood sample, including arterial, venous, or capillary) with a standard laboratory method (usually glucose oxidase or hexokinase methods). We excluded continuous glucose monitoring devices, as these have been addressed in a recent Cochrane review.[20] Studies were not excluded if the laboratory comparison method was unclear or deemed inappropriate; instead, these concerns were addressed in quality assessment.

The focus population for this review was neonates. Studies that did not provide sufficient information for a 2 × 2 contingency table for meta-analysis inclusion were included in a summary table of device performance.

Data Extraction

Four review authors independently conducted abstract and title screening, followed by full-text review using the Covidence tool,[21] to assess study eligibility. Subsequently, 2 reviewers independently extracted data from the retrieved studies using a prespecified data extraction form. Any discrepancies at any stage were resolved through discussion or with a third review author. The extracted data included study details such as publication date, study design, setting, definition of target condition, testing devices utilized, point-of-care test employed, and outcomes. For included studies that provided sufficient data, we constructed contingency tables comparing the point-of-care test with the reference standard. For studies that provided insufficient data or addressed a nontarget population, any numeric data (mean difference, mean bias, correlation coefficient) were extracted into a table alongside study details and the author's conclusions.

Quality Assessment

Two reviewers assessed the quality of studies included in meta-analyses using the recommended Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.[22] Disagreements were resolved through discussion or with another review author. The assessed domains include participant selection, index test (point-of-care test), reference standard, flow, and timing. All domains were evaluated for risk of bias, and the first 3 domains for applicability concerns. The quality assessment summaries of meta-analyzed studies are presented in tables generated using Review Manager (RevMan) 5.4.1 Software (The Cochrane Collaboration).[23]

Data Synthesis and Analysis

The unit of analysis was the sample that was tested for neonatal hypoglycemia. The contingency tables from included studies were used to calculate the sensitivity and specificity of each point-of-care device methodology. The definition of hypoglycemia was as defined by the study. Findings were presented graphically through estimate plots in a forest plot constructed using Review Manager 5.4.1[23] to allow visual assessment of test accuracy. We grouped studies based on their device methodology (reaction enzyme used [hexokinase (HK), glucose oxidase (GO), or glucose-1-dehydrogenase (GDH)], and photometric or electrochemical measurement). Summary receiver operating characteristic (SROC) curves were constructed for each group. A hierarchical random-effects model, which accounts for unexplained heterogeneity between studies, was employed using STATA 17[24] to generate pooled results. Analyses were classified as quantitative if they included sufficient data to calculate true positives, false positives, true negatives, and false negatives, enabling the formation of a 2 × 2 contingency table.

Studies lacking sufficient data for 2 × 2 contingency tables were classified as qualitative, and their findings were summarized in a table indicating the directionality of device performance (underestimation, good estimation, and overestimation). The original Clarke error grid considers values within 20% of the reference sensor to be acceptable,[25] but in the context of neonatal hypoglycemia, this margin was considered too wide because of the potentially clinically significant implications of missed cases or unnecessary treatment due to under- or overestimation. We used a 10% margin, based on the widely accepted hypoglycemia threshold of 2.6 mmol/L,[26-28] to inform our definitions of underestimation and overestimation. Specifically, we

considered any measured glucose concentrations more than 0.26 mmol/L below the threshold (ie, <2.34 mmol/L) as an underestimation and any measured glucose concentrations more than 0.26 mmol/L above the threshold (ie, >2.86 mmol/L) as an overestimation. These thresholds were applied to assess the accuracy of glucose measurements in relation to the clinical definition of hypoglycemia. Studies lacking numerical data (mean difference or bias) were categorized depending on the authors' conclusions or other presented results.

We assessed certainty of evidence for sensitivity and specificity using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach[29] and created a "Summary of Findings" table using the Grade Pro Guideline Development Tool.[30]

Results

Study Selection

The initial search yielded 11 661 records, of which 8034 records underwent title and abstract screening after removing duplicates. Among these, 549 records were selected for full-text review, resulting in 71 eligible studies (comprising 73 records). Of these, 31 studies were included in the quantitative analysis, while 40 were included in the qualitative analysis (Figure 1).

Study Characteristics

Among the 31 studies included in the quantitative analysis, a total of 6756 patients and 19 119 samples were assessed.

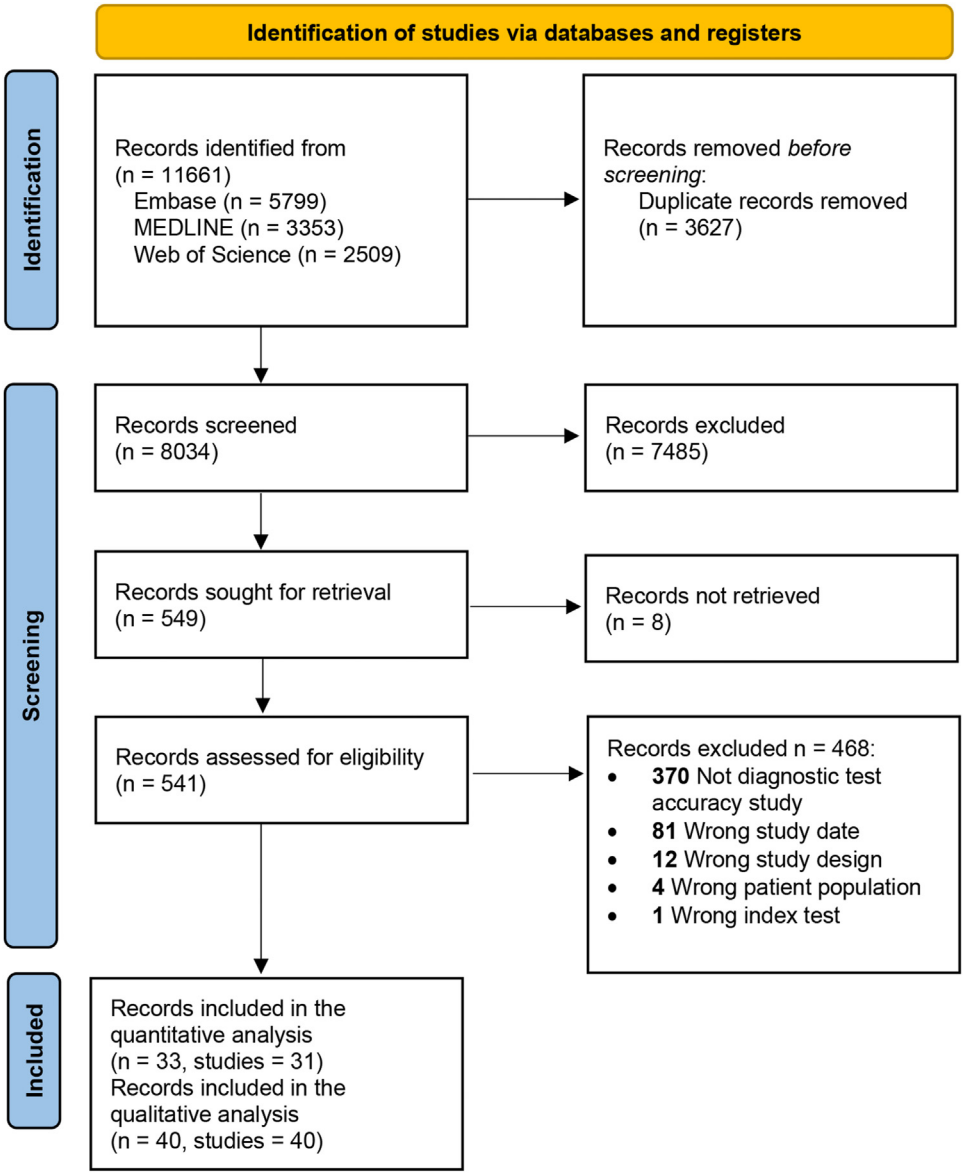

Figure 1. Flow diagram of the included studies.

Seventeen studies were conducted in high-income countries, 10 in upper-middle-income countries, and 4 in lower-middle-income countries. Eight studies (10 devices) evaluated point-of-care testing using GO + photometry, 6 studies (6 devices) used GDH + photometry, 13 studies (14 devices) used GO + electrochemistry, 12 studies (12 devices) used GDH + electrochemistry, and 2 studies (2 devices) used HK + electrochemistry. All these studies included neonatal populations and were cross-sectional. The general characteristics of the eligible studies are described in **Appendix 3** (available at www.jpeds.com).

Risk of Bias

Six of the 31 studies (19.4%) were at high risk of patient selection bias, indicating that participants were enrolled based on convenience rather than random or consecutive sampling (**Figure 2**). Five of the 31 studies (16.1%) were at high risk of bias related to the point-of-care test, meaning that the conduct or interpretation of the point-of-care test introduced bias. Three studies (9.7%) were at high risk of bias for the reference standard because the reference test was conducted with knowledge of the point-of-care test result. Two studies (6.5%) were at high risk of bias due to flow and timing issues, such as inappropriate intervals between the point-of-care and reference tests or unclear participant exclusion criteria. In addition, 7 of the 31 studies (22.6%) had high risk regarding the applicability of patient selection.

Analysis Results

Diagnostic accuracy showed significant variation, with sensitivity ranging from 25% to 100% and specificity from 7% to 100% in individual studies (**Figure 3**). For GO + photometry methods, the mean sensitivity was 0.72 (95% CI: 0.64-0.76, moderate certainty), and the mean specificity was 0.95 (95% CI: 0.87-0.98, low certainty) (**Table I**). For GHD + photometry methods, the mean sensitivity was 0.64 (95% CI: 0.13-0.95, very low certainty), and the mean specificity was 0.99 (95% CI: 0.88-1.00, high certainty). For GO + electrochemistry methods, the pooled sensitivity was 0.82 (95% CI: 0.70-0.89, moderate certainty), and pooled specificity was 0.94 (95% CI: 0.83-0.98. moderate

Figure 2. Risk of bias and applicability concerns. **A.** Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies. **B.** Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

GO + photometry

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Ellis 1996	26	23	5	39	2.0	0.84 [0.66, 0.95]	0.63 [0.50, 0.75]
Falcao 1997	8	18	6	432	1.9	0.57 [0.29, 0.82]	0.96 [0.94, 0.98]
Giep 1996	9	4	2	171	2.2	0.82 [0.48, 0.98]	0.98 [0.94, 0.99]
Giep 1996	9	1	1	171	2.2	0.90 [0.55, 1.00]	0.99 [0.97, 1.00]
Ho 1991	15	35	2	154	2.2	0.88 [0.64, 0.99]	0.81 [0.75, 0.87]
Ho 1991	14	19	3	170	2.2	0.82 [0.57, 0.96]	0.90 [0.85, 0.94]
Kitsommart 2013	70	8	25	74	2.2	0.74 [0.64, 0.82]	0.90 [0.82, 0.96]
Meloy 1999	53	22	17	156	2.2	0.76 [0.64, 0.85]	0.88 [0.82, 0.92]
Ngerncham 2012	49	0	43	88	2.2	0.53 [0.43, 0.64]	1.00 [0.96, 1.00]
Raizman 2016	14	28	8	1622	2.1	0.64 [0.41, 0.83]	0.98 [0.98, 0.99]

GDH + photometry

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Belini 2007	2	0	6	69	2.2	0.25 [0.03, 0.65]	1.00 [0.95, 1.00]
Deshpande 1996	11	0	4	100	2.6	0.73 [0.45, 0.92]	1.00 [0.96, 1.00]
Ellis 1996	0	0	31	62	2.0	0.00 [0.00, 0.11]	1.00 [0.94, 1.00]
Hamid 2004	99	13	2	178	2.2	0.98 [0.93, 1.00]	0.93 [0.89, 0.96]
Nayeri 2014	20	4	5	190	2.5	0.80 [0.59, 0.93]	0.98 [0.95, 0.99]
Nooripoor 2012	55	18	3	80	2.5	0.95 [0.86, 0.99]	0.82 [0.73, 0.89]

GO + electrochemistry

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Balion 2006	13	9	12	415	2.6	0.52 [0.31, 0.72]	0.98 [0.96, 0.99]
Diaw 2013	47	56	8	149	2.5	0.85 [0.73, 0.94]	0.73 [0.66, 0.79]
Kitsommart 2013	56	3	39	74	2.2	0.59 [0.48, 0.69]	0.96 [0.89, 0.99]
Makaya 2012	9	16	1	702	2.0	0.90 [0.55, 1.00]	0.98 [0.96, 0.99]
Michel 2005	29	15	17	808	2.6	0.63 [0.48, 0.77]	0.98 [0.97, 0.99]
Nasiri 2016	153	24	7	20	2.2	0.96 [0.91, 0.98]	0.45 [0.30, 0.61]
Ngerncham 2012	57	0	35	88	2.2	0.62 [0.51, 0.72]	1.00 [0.96, 1.00]
Nuntnarumit 2011	20	0	1	130	2.5	0.95 [0.76, 1.00]	1.00 [0.97, 1.00]
Raizman 2016	10	16	4	1250	2.1	0.71 [0.42, 0.92]	0.99 [0.98, 0.99]
Roth-Kleiner 2010	122	66	20	264	2.5	0.86 [0.79, 0.91]	0.80 [0.75, 0.84]
Roth-Kleiner 2010	33	0	27	260	2.5	0.55 [0.42, 0.68]	1.00 [0.99, 1.00]
Tendl 2013	9	3	0	147	2.5	1.00 [0.66, 1.00]	0.98 [0.94, 1.00]
Torkaman 2016	44	180	1	13	2.6	0.98 [0.88, 1.00]	0.07 [0.04, 0.11]
Wang 2013	41	16	4	246	2.6	0.91 [0.79, 0.98]	0.94 [0.90, 0.96]

GDH + electrochemistry

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Diaw 2013	73	52	9	140	2.5	0.89 [0.80, 0.95]	0.73 [0.66, 0.79]
Diaw 2013	49	75	6	122	2.5	0.89 [0.78, 0.96]	0.62 [0.55, 0.69]
Dietzen 2015	24	3	0	606	2.2	1.00 [0.86, 1.00]	1.00 [0.99, 1.00]
Foster 2013	50	3	7	30	2.6	0.88 [0.76, 0.95]	0.91 [0.76, 0.98]
Gallagher 2000	17	2	0	37	2.6	1.00 [0.80, 1.00]	0.95 [0.83, 0.99]
Ghergherehchi 2015	30	34	35	101	2.5	0.46 [0.34, 0.59]	0.75 [0.67, 0.82]
Harish 2015	43	9	60	138	2.5	0.42 [0.32, 0.52]	0.94 [0.89, 0.97]
Kumar 2015	121	1	57	321	2.2	0.68 [0.61, 0.75]	1.00 [0.98, 1.00]
Nuntnarumit 2011	12	0	9	130	2.5	0.57 [0.34, 0.78]	1.00 [0.97, 1.00]
Raff 2012	485	46	17	69	2.5	0.97 [0.95, 0.98]	0.60 [0.50, 0.69]
Roth-Kleiner 2010	62	8	82	380	2.5	0.43 [0.35, 0.52]	0.98 [0.96, 0.99]
Wang 2013	35	3	10	259	2.6	0.78 [0.63, 0.89]	0.99 [0.97, 1.00]

HK + electrochemistry

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Gallagher 2000	17	4	0	35	2.6	1.00 [0.80, 1.00]	0.90 [0.76, 0.97]
Janjindamai 2022	39	5	11	90	2.6	0.78 [0.64, 0.88]	0.95 [0.88, 0.98]

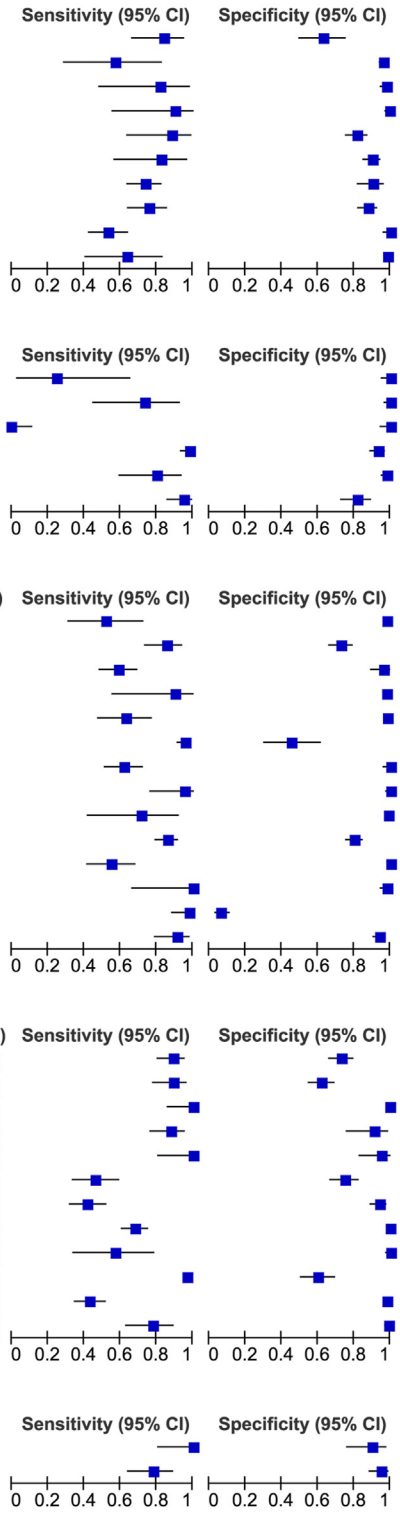

Figure 3. Forest plot of sensitivity and specificity with 95% confidence interval for detection of hypoglycaemia in each study by different testing methods.

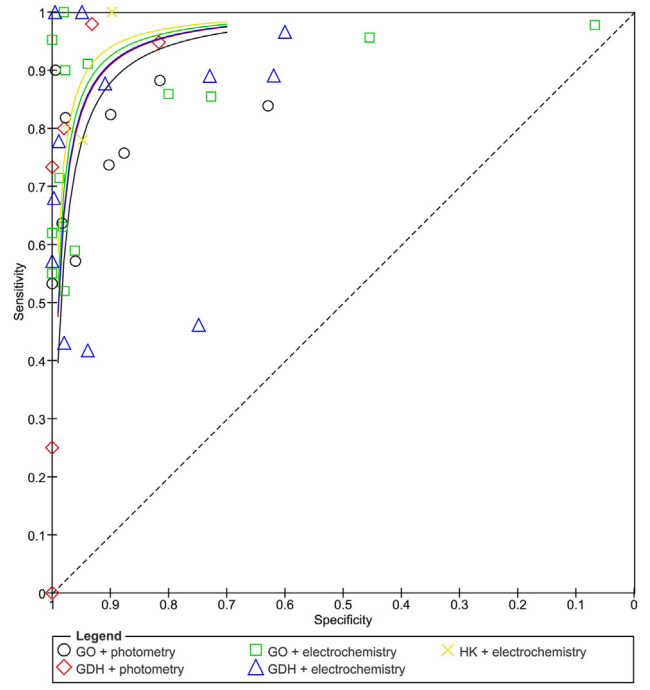

Figure 4. Summary receiver operating characteristic curves (SROC) showing the discriminant ability of point-of-care testing methods for different testing methods.

certainty). For GDH + electrochemistry methods, the pooled sensitivity was 0.81 (95% CI: 0.62-0.91, very low certainty), and pooled specificity was 0.96 (95% CI: 0.88-0.99, low certainty). For HK + electrochemistry methods, the pooled sensitivity was 0.84 (95% CI: 0.73-0.91, moderate certainty), and pooled specificity was 0.93 (95% CI: 0.88-0.96, moderate certainty) (Figure 4).

Summary of Findings

The GRADE assessment showed that the certainty of sensitivity for different methods ranged from very low to moderate. This variability was due to downgrading for factors such as unclear overall risk of bias, significant heterogeneity, or wide confidence intervals. Similarly, the certainty of specificity for different testing methods ranged from low to high, influenced by downgrading for unclear overall risk of bias or significant heterogeneity (Table II).

Studies Included in Qualitative Analysis

Forty studies were only included in the qualitative analysis due to a lack of sufficient information to generate the contingency table. For GO + photometry, most studies reported good estimates (42.1%) or underestimates (36.8%), with fewer studies reporting overestimates (21.1%). In contrast, for GDH + photometry, over half of the studies reported overestimates (50%), with the remaining studies showing good estimates (40%) or underestimates (10%). When using GO + electrochemistry, the studies were evenly split between good estimates (42.9%) and overestimates (42.9%), with fewer underestimates (14.3%). Finally, for GDH + electrochemistry, half of the studies reported good estimates (50%), and the rest were divided between overestimates (33.3%) and underestimates (16.7%) (Table III).

Discussion

We conducted a systematic review and meta-analysis to ascertain the accuracy of different point-of-care methods for measuring blood glucose concentrations in neonates. The variability in diagnostic accuracy observed across different methods highlights the challenges in achieving consistent and reliable glucose measurements for this group. GO + photometry and GDH + photometry methods show high specificity but low sensitivities, with a tendency to overestimate hypoglycemia. GDH + electrochemistry methods provide balanced performance with low to very low certainty evidence, while GO + electrochemistry and HK + electrochemistry methods demonstrated consistent performance with moderate certainty evidence. Overall, the persistent issues of underestimation and overestimation across all these methods emphasize ongoing challenges in achieving reliable point-of-care blood glucose measurements in neonates.

Factors contributing to the variability in diagnostic accuracy across methods could include differences in study design, sample handling, calibration protocols, and inherent limitations of each testing method. Variations in study design, such as sample sizes, study populations, and the clinical settings in which tests are performed, can significantly impact the results. For instance, studies with small sample sizes may not adequately represent the broader population, leading to variability in performance metrics like sensitivity

Table I. Pooled mean sensitivity and specificity of point-of-care device methodology for detection of hypoglycemia

Point-of-care test (number of studies)	Cases*/Samples (%)	Sensitivity (95% CI)	I^2 (%) for sensitivity	Specificity (95% CI)	I^2 (%) for specificity
GO + photometry (devices = 10, studies = 8)	379/3614 (10.5)	0.72 (0.64, 0.76)	35.8	0.95 (0.87, 0.98)	81.1
GDH + photometry (devices = 6, studies = 6)	238/952 (25)	0.64 (0.13, 0.95)	51.2	0.99 (0.88, 1.00)	3.0
GO + electrochemistry (devices = 14, studies = 13)	819/5791 (14.1)	0.82 (0.70, 0.89)	73.9	0.94 (0.83, 0.98)	82.6
GDH + electrochemistry (devices = 12, studies = 11)	1293/3862 (33.5)	0.81 (0.62, 0.91)	85.9	0.96 (0.88, 0.99)	69.0
Hexokinase + electrochemistry (devices = 2, studies = 2)	67/201 (33.3)	0.84 (0.73, 0.91)	NA	0.93 (0.88, 0.96)	NA

*Number of samples below the threshold for diagnosis of hypoglycemia as defined by each study.

Table II. Summary of findings table

Patient or population: Neonates; setting: any birth settings; reference test: standard laboratory method

Test result	Number of results per 1000 patients tested (95% CI) Prevalence 50% Typically seen in at-risk babies	Number of participants (studies)	Certainty of the evidence (GRADE)
Enzymatic (GO) + photometry: Pooled sensitivity:0.72 (95% CI: 0.64-0.76) | Pooled specificity: 0.95 (95% CI: 0.87-0.98)			
True positives	360 (320-380)	3614 (10)	⨁⨁⨁◯ Moderate*
False negatives	140 (120-180)		
True negatives	475 (435-490)	3614 (10)	⨁⨁◯◯ Low*,†
False positives	25 (10-65)		
Enzymatic (GDH) + photometry: Pooled sensitivity: 0.64 (95% CI: 0.13-0.95) | Pooled specificity: 0.99 (95% CI: 0.88-1.00)			
True positives	320 (65-475)	952 (6)	⨁◯◯◯ Very low†,‡
False negatives	180 (25-435)		
True negatives	495 (440-500)	952 (6)	⨁⨁⨁⨁ High
False positives	5 (0-60)		
Enzymatic (GO) + electrochemistry: Pooled sensitivity: 0.82 (95% CI: 0.70-0.89) | Pooled specificity: 0.94 (95% CI: 0.83-0.98)			
True positives	410 (350-445)	5791 (14)	⨁⨁⨁◯ Moderate†
False negatives	90 (55-150)		
True negatives	470 (415-490)	5791 (14)	⨁⨁⨁◯ Moderate†
False positives	30 (10-85)		
Enzymatic (GDH) + electrochemistry: Pooled sensitivity: 0.81 (95% CI: 0.62-0.91) | Pooled specificity: 0.96 (95% CI: 0.88-0.99)			
True positives	405 (310-455)	3862 (12)	⨁◯◯◯ Very low*,†,§
False negatives	95 (45-190)		
True negatives	480 (440-495)	3862 (12)	⨁⨁◯◯ Low*,†
False positives	20 (5-60)		
Enzymatic (hexokinase): Pooled sensitivity: 0.84 (95% CI: 0.73-0.91) | Pooled specificity: 0.93 (95% CI: 0.88-0.96)			
True positives	420 (365-455)	201 (2)	⨁⨁⨁◯ Moderate*
False negatives	80 (45-135)		
True negatives	465 (440-480)	201 (2)	⨁⨁⨁◯ Moderate*
False positives	35 (20-60)		

*Downgraded one level of risk of bias due to overall unclear risk of bias.
†Downgraded one level for inconsistency due to significant heterogeneity.
‡Downgraded 2 levels for imprecision due to the very wide range of confidence intervals.
§Downgraded one level for imprecision due to the wide range of CIs.

and specificity. In addition, differences in how hypoglycemia is defined and diagnosed across studies can affect the accuracy of test performance assessments. For example, studies using lower glucose thresholds to define hypoglycemia may increase the risk of false negatives, which is especially concerning given the severe implications of missed neonatal hypoglycemia. Conversely, studies using higher or multiple thresholds may lead to discrepancies in identifying true positive or negative cases, potentially leading to unnecessary interventions. The effect of variability in diagnostic thresholds underscores the critical importance of selecting appropriate point-of-care devices tailored to clinical settings to ensure reliable identification of hypoglycemia.

Proper sample handling is also crucial for accurate glucose measurement. Variability in the collection, processing, and storage of blood samples can lead to inconsistencies in glucose measurement results. Differences in how samples are collected (eg, type of device or technique), processed (eg, timing and handling procedures), and stored (eg, temperature and duration) can all affect the accuracy and reliability of the test outcomes.[31] Each glucose measurement method also has its own limitations. Enzymatic methods may have varying sensitivities to interference from other substances in the blood, such as proteins, lipids, or medications, that can mimic the glucose signal or inhibit enzyme activity, leading to inaccurate glucose readings.[11] Photometric methods may be influenced by light absorption properties of various substances in the blood, such as hemoglobin or other pigments, that alter the amount of light absorbed or transmitted through the sample, impacting the reliability of the test results.[32] Electrochemical methods can be affected by electrode stability and the electrochemical environment.[33] These inherent limitations can lead to differences in how accurately each method detects neonatal hypoglycemia, affecting overall test reliability. In addition, variability across different device manufacturers using the same methods may also influence performance.

Because the clinical implications of failing to detect cases of neonatal hypoglycemia can be severe, including neurodevelopmental impairment,[4,5] the sensitivity of point-of-care tests may be more important than specificity in this population, as false negative rates impact more on groups where the prevalence of a condition is higher.[34] Across the 31 studies, all testing methods demonstrated high specificity (≥93%), whereas sensitivity was highest in studies using HK/GDH/GO + electrochemistry, suggesting that these methods have a lower risk of false negatives that may lead to unnecessary treatment.[35] The SROC curve confirmed the superior

Table III. Directionality of glucose measurement estimates relative to a 10% margin for hypoglycemia

Device method	Underestimates	Good estimate	Overestimates
GO + photometry (n = 19)	n = 7 Reynolds 1993 (BM-Reflolux) Reynolds 1994 (BM-Reflolux) LouderIll 1996 (Accu-Chek II) LouderIll 1996 (One-Touch) Schlebusch 1998 (One-Touch) Ho 2004 (BM-Reflolux) Duke 2022 (Vitros)	n = 8 Schlebusch 1993 (BM-Reflolux) Kirkham 1995 (BM-Reflolux) Garland 1996 (One-Touch) Innanen 1997 (Ames Glucometer) Schlebusch 1998 (Accu-Chek III) St-Louis 2002 (SureStep) Kiattimongkol 2003 (Medisafe Mini) Wada 2015 (Medisafe Mini) 8	n = 4 Kirkham 1995 (Ames Glucometer) Sharief 1997 (BM-Reflolux) Papp 2001 (BM-Reflolux) Sudha Reddy 2014 (B Braun)
GDH + photometry (n = 10)	n = 1 Dahlberg 1997 (HemoCue)	n = 4 Schlebusch 1998 (HemoCue) Raile 1998 (HemoCue) Warner 2011 (HemoCue) Warner 2011 (Optium)	n = 5 Leonard 1997 (HemoCue) Sharief 1997 (HemoCue) Ziljstra 1998 (HemoCue) Upadrasta 2012 (HemoCue) Sudha Reddy 2014 (HemoCue)
GO + electrochemistry (n = 14)	n = 2 Demers 1999 (Precision-G) Fokkert 2012 (StatStrip)	n = 6 Sonderkaer 1999 (ABL 625) Newman 2002 (AVL Omni 9) Peet 2002 (B860) St-Louis 2002 (Precision PCx) Ho 2004 (Elite XL) Ho 2004 (Precision)	n = 6 Schlebusch 1998 (Glucometer Elite) Thomas 2000 (Precision-G) McNamara 2001 (EML 105) Upadrasta 2012 (Seimens 1265) Foster 2013 (StatStrip) Wada 2015 (StatStrip)
GDH + electrochemistry (n = 18)	n = 3 Kim 2021 (PRO) Kim 2021 (H Beat) Brooks 2023 (Inform II)	n = 9 Thomas 2000 (Advantage) St-Louis 2002 (Inform) Rosenthal 2006 (Accutrend) Sievenpiper 2011 (Precision PXP) Warner 2011 (Performa) Dietzen 2013 (Inform II) Dietzen 2013 (Performa) Dietzen 2013 (Aviva) Zayek 2019 (Inform II) Kim 2021 (Inform II) Kim 2021 (BAROzen)	n = 6 McNamara 2001 (Advantage) Papp 2001 (Advantage) Ho 2004 (Advantage) Foster 2013 (Performa) BenAmeur 2016 (Active) BenAmeur 2016 (Performa)
Other methods		Gong 2012 (hexose on DBS)	Dixon 2023 (light-based sensor)
Mechanism not described or unable to determine*	Woods 2002 (Glucotrend)	Ho 2004 (Glucotrend)	BenAmeur 2016 (Bionime)

GO, glucose oxidase; *GDH*, glucose dehydrogenase.
Table lists study (instrument tested).
*Testing method not explained in sufficient detail in paper (method, manufacturer or model not specified).

discriminative ability of the HK + electrochemistry over the GDH + electrochemistry and GO + electrochemistry methods to maximize the detection of neonatal hypoglycemia whilst minimizing false positives.[36]

The SROC curve suggests that the GO + photometry and GDH + photometry methods are the least accurate methods examined in the meta-analysis. This is consistent with our qualitative analysis showing that the GO + photometry method underestimated the incidence of neonatal hypoglycemia (ie, missing more than 10% of true cases) in over a third of the included studies, while the GDH + photometry method overestimated hypoglycemia (ie, more than 10% of identified cases were false positives) in half of the included studies. However, the qualitative analysis also showed that of the 3 methods that appeared most accurate on meta-analysis, GDH + electrochemistry over-estimated the incidence of hypoglycemia in 6 of 18 studies (33%), and GO + electrochemistry in 6 of 14 studies (43%).

The HK + electrochemistry method demonstrated consistent performance with moderate certainty evidence. This method had high sensitivity, which is particularly beneficial for accurately detecting neonatal hypoglycemia, a condition where false negatives can have severe clinical implications. However, the limited number of studies suggests the need for further research to confirm these findings across different devices, populations and diagnostic thresholds. In addition to diagnostic accuracy, the feasibility and cost-effectiveness of the HK + electrochemistry method are also critical considerations, particularly in low-resource environments. Implementation of this method may be limited by higher device costs, calibration requirements, and infrastructure limitations. Overall, the HK + electrochemistry approach represents a promising technique for neonatal glucose monitoring, but further validation and refinement are needed to maximize its diagnostic accuracy.

Although there appear to be high levels of heterogeneity (ie, ≥75%) for specificity and sensitivity across devices using the GDH + electrochemistry, and GO + electrochemistry, the interpretation of I^2 measures is not straightforward,[37] and, due to the threshold effect, arguably not well-suited to the evaluation of heterogeneity of specificity and sensitivity of diagnostic tests in meta-analyses.[38] There are also likely variations in study design and execution across studies, each of

which can introduce heterogeneity in results.[39] However, we employed a hierarchical random-effects model, thus accounting for any unexplained heterogeneity between studies included in our meta-analysis.[39]

The systematic review and meta-analysis have some limitations. One significant challenge was the inability to perform subgroup analysis based on different diagnostic thresholds. This subgroup analysis was planned to assess how varying cutoff values impact device accuracy, but the categorization of studies by different methods and thresholds resulted in insufficient data for meaningful comparisons, particularly as there was little variation in the thresholds used in studies grouped by the different testing methods. Furthermore, the included studies exhibited considerable heterogeneity in design, methodology, and patient populations, complicating the synthesis of results and affecting the generalizability of the findings. Further, the limited number of studies evaluating certain methods, particularly HK + electrochemistry, highlights the need for additional studies. Such studies should be multicentered, incorporate standardized study designs, and pool data across diverse populations to enhance the robustness and generalizability of findings. Variability in testing conditions and calibration practices further contributed to the inconsistency, as differing sample handling and calibration procedures were not uniformly reported across studies. These limitations underscore the need for additional research to address these gaps and enhance the assessment of reliability of point-of-care devices for the detection of neonatal hypoglycemia.

The applicability of these results may be limited by the fact that clinicians working in healthcare facilities do not have access to the most accurate point-of-care testing devices. Furthermore, implementing these accurate point-of-care testing devices in developing countries will inherently be more challenging, owing to factors such as healthcare infrastructure and cost.[40] However, it is hoped that these results will be useful to those responsible for purchasing new equipment when updating is practicable.

In summary, meta-analysis and subsequent analysis revealed that 3 device methods had greater sensitivity for detecting neonatal hypoglycemia: HK + electrochemistry, GDH + electrochemistry and GO + electrochemistry. Among these, the HK + electrochemistry method was the most accurate in detecting hypoglycemia within the neonatal population, although there were a limited number of studies using this method. ■

CRediT authorship contribution statement

Sophie L. St Clair: Writing – review & editing, Software, Methodology, Data curation. **Caitlyn M. Ulyatt:** Writing – review & editing, Validation, Software, Methodology, Data curation. **Maria T. Corkin:** Writing – review & editing, Validation, Software, Data curation. **Libby G. Lord:** Writing – review & editing, Validation, Software, Data curation. **Caroline A. Crowther:** Writing – review & editing, Visualization, Validation, Supervision, Funding acquisition, Conceptualization. **Jane E. Harding:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Luling Lin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors are funded partly by the Health Research Council of New Zealand (J.H., C.C.), the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (M.C., L.G.L., S.S.), the Aotearoa Foundation (L.L.), and the New Zealand Ministry of Business, Innovation, and Employment (C.U.). All authors reviewed and approved the final manuscript as submitted and agree to be accountable for all aspects of the work. The authors declare no competing interests.

Submitted for publication Oct 7, 2024; last revision received Nov 27, 2024; accepted Dec 10, 2024.

Reprint requests: Luling Lin, PhD, Liggins Institute, 85 Park Road, Grafton, Auckland 1023, New Zealand. E-mail: Luling.lin@auckland.ac.nz

References

1. Hay WW Jr, Raju TN, Higgins RD, Kalhan SC, Devaskar SU. Knowledge gaps and research needs for understanding and treating neonatal hypoglycemia: workshop report from eunice kennedy shriver national institute of child health and human development. J Pediatr 2009;155:612-7.
2. Harris DL, Weston PJ, Harding JE. Incidence of neonatal hypoglycemia in babies identified as at risk. J Pediatr 2012;161:787-91.
3. Adamkin DH. Postnatal glucose homeostasis in late-preterm and term infants. Pediatrics 2011;127:575-9.
4. Kaiser JR, Bai S, Gibson N, Holland G, Lin TM, Swearingen CJ, et al. Association between transient newborn hypoglycemia and fourth-grade achievement test proficiency: a population-based study. JAMA Pediatr 2015;169:913-21.
5. Lucas A, Morley R, Cole TJ. Adverse neurodevelopmental outcome of moderate neonatal hypoglycaemia. BMJ 1988;297:1304-8.
6. Adamkin DH. Neonatal hypoglycemia. Curr Opin Pediatr 2016;28:150-5.
7. Giouleka S, Gkiouleka M, Tsakiridis I, Daniilidou A, Mamopoulos A, Athanasiadis A, et al. Diagnosis and management of neonatal hypoglycemia: a comprehensive review of guidelines. Children 2023;10:1-22.
8. Galderisi A, Facchinetti A, Steil GM, Ortiz-Rubio P, Cavallin F, Tamborlane WV, et al. Continuous glucose monitoring in very preterm infants: a randomized controlled trial. Pediatrics 2017;140:e20171162.
9. Beardsall K. Measurement of glucose levels in the newborn. Early Hum Dev 2010;86:263-7.
10. Sirkin A, Jalloh T, Lee L. Selecting an accurate point-of-care testing system: clinical and technical issues and implications in neonatal blood glucose monitoring. J Pediatr Nurs 2002;7:104-12.
11. Erbach M, Freckmann G, Hinzmann R, Kulzer B, Ziegler R, Heinemann L, et al. Interferences and limitations in blood glucose self-testing: an overview of the current knowledge. J Diabetes Sci Technol 2016;10:1161-8.
12. Vashist SK, Luppa PB, Yeo LY, Ozcan A, Luong JHT. Emerging technologies for next-generation point-of-care testing. Trends Biotechnol 2015;33:692-705.

13. Harding JE, Harris DL, Hegarty JE, Alsweiler JM, McKinlay CJ. An emerging evidence base for the management of neonatal hypoglycaemia. Early Hum Dev 2017;104:51-6.
14. Woo HC, Tolosa L, El-Metwally D, Viscardi RM. Glucose monitoring in neonates: need for accurate and non-invasive methods. Arch Dis Child Fetal Neonatal Ed 2014;99:F153-7.
15. Bellini C, Serra G, Risso D, Mazzella M, Bonioli E. Reliability assessment of glucose measurement by HemoCue analyser in a neonatal intensive care unit. Clin Chem Lab Med 2007;45:1549-54.
16. Ortiz-Martínez M, Flores-Delatoba R, González-González M, Rito-Palomares M. Current challenges and future trends of enzymatic paper-based point-of-care testing for diabetes mellitus type 2. Biosensors 2021;11:1-22.
17. Parker K, Lyon ME, Kyle BD, Strueby L, Inman M. The clinical effect of glucose meter selection upon the detection of neonatal hypoglycemia. Paediatr Child Health 2021;27:12-4.
18. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy 2.0 (updated July 2023) [Internet]: Cochrane 2023. Accessed December 15, 2023. https://training.cochrane.org/handbook-diagnostic-test-accuracy/current
19. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, et al., the P-DTAG. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 2018;319:388-96.
20. Galderisi A, Trevisanuto D, Russo C, Hall R, Bruschettini M. Continuous glucose monitoring for the prevention of morbidity and mortality in preterm infants. Cochrane Database Syst Rev 2021;12:CD013309.
21. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia. Accessed January 12, 2024. www.covidence.org
22. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.
23. Review manager (RevMan) [Computer program]. 5.4.1 ed. Copenhagen: The Cochrane Collaboration; 2020.
24. Stata Statistical Software: Release 17. StataCorp. College Station, TX: StataCorp LLC; 2021.
25. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10:622-8.
26. Harris DL, Weston PJ, Battin MR, Harding JE. A survey of the management of neonatal hypoglycaemia within the Australian and New Zealand Neonatal Network. J Paediatr Child Health 2014;50:E55-62.
27. Dixon KC, Ferris RL, Marikar D, Chong M, Mittal A, Manikam L, et al. Definition and monitoring of neonatal hypoglycaemia: a nationwide survey of NHS England neonatal units. Arch Dis Child Fetal Neonatal Ed 2017;102:F92-3.
28. Clinical Practice Committee. Guidelines for the management of hypoglycaemia. 2018. Accessed August 29, 2024. http://www.adhb.govt.nz/newborn/Guidelines/Nutrition/HypoglycaemiaManagement.htm
29. Schünemann H, Brożek J, Guyatt G, Oxman A, eds. GRADE handbook for grading quality of evidence and strength of recommendations. Hamilton, Ontario, Canada: McMaster University; 2013. The GRADE Working Group. Accessed December 30, 2024. https://gdt.gradepro.org/app/handbook/handbook.html
30. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. Hamilton, Ontario, Canada: McMaster University; Kraków, Poland: Evidence Prime. 2024. Accessed December 30, 2024. https://www.gradepro.org/
31. Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. Anal Bioanal Chem 2015;407:4879-92.
32. Villena GW, Mobashsher AT, Abbosh A. The progress of glucose monitoring-a review of invasive to minimally and non-invasive techniques, devices and sensors. Sensors 2019;19:800.
33. Li W, Luo W, Li M, Chen L, Chen L, Guan H, et al. The impact of recent developments in electrochemical POC sensor for blood sugar care. Front Chem 2021;9:723186.
34. Shivkumar S, Peeling R, Jafari Y, Joseph L, Pai NP. Accuracy of rapid and point-of-care screening tests for Hepatitis C: a systematic review and meta-analysis. Ann Intern Med 2012;157:558-66.
35. Bentley TG, Catanzaro A, Ganiats TG. Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis. BMC Res Notes 2015;5:1-7.
36. Walter SD. The partial area under the summary ROC curve. Stat Med 2005;24:2025-40.
37. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I(2) is not an absolute measure of heterogeneity. Res Synth Methods 2017;8:5-18.
38. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. Korean J Radiol 2015;16:1188-96.
39. Langan D. Assessing heterogeneity: heterogeneity in random-effects meta-analysis. In: Evangelou E, Veroniki AA, eds. Meta-research: methods and protocols. New York, NY: Springer US; 2022. p. 67-89.
40. Mitra P, Sharma P, Praveen S. POCT in developing countries. EJIFCC 2021;32:195-9.