

University of Groningen

## Diagnostic Accuracy of Portable, Handheld Point-of-Care Tests vs Laboratory-Based Bilirubin Quantification in Neonates

Westenberg, Lauren E.H.; Been, Jasper V.; Willemsen, Sten P.; Vis, Jolande Y.; Tintu, Andrei N.; Bramer, Wichor M.; Dijk, Peter H.; Steegers, Eric A.P.; Reiss, Irwin K.M.; Hulzebos, Christian V.

*Published in:*  
 JAMA Pediatrics

*DOI:*  
[10.1001/jamapediatrics.2023.0059](https://doi.org/10.1001/jamapediatrics.2023.0059)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
 Publisher's PDF, also known as Version of record

*Publication date:*  
 2023

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Westenberg, L. E. H., Been, J. V., Willemsen, S. P., Vis, J. Y., Tintu, A. N., Bramer, W. M., Dijk, P. H., Steegers, E. A. P., Reiss, I. K. M., & Hulzebos, C. V. (2023). Diagnostic Accuracy of Portable, Handheld Point-of-Care Tests vs Laboratory-Based Bilirubin Quantification in Neonates: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, 177(5), 479-488. <https://doi.org/10.1001/jamapediatrics.2023.0059>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Diagnostic Accuracy of Portable, Handheld Point-of-Care Tests vs Laboratory-Based Bilirubin Quantification in Neonates

## A Systematic Review and Meta-analysis

Lauren E. H. Westenberg, MD; Jasper V. Been, MD, PhD; Sten P. Willemsen, PhD; Jolande Y. Vis, MD, PhD; Andrei N. Tintu, PhD; Wichor M. Bramer, PhD; Peter H. Dijk, MD, PhD; Eric A. P. Steegers, MD, PhD; Irwin K. M. Reiss, MD, PhD; Christian V. Hulzebos, MD, PhD

[+ Supplemental content](#)

**IMPORTANCE** Quantification of bilirubin in blood is essential for early diagnosis and timely treatment of neonatal hyperbilirubinemia. Handheld point-of-care (POC) devices may overcome the current issues with conventional laboratory-based bilirubin (LBB) quantification.

**OBJECTIVE** To systematically evaluate the reported diagnostic accuracy of POC devices compared with LBB quantification.

**DATA SOURCES** A systematic literature search was conducted in 6 electronic databases (Ovid MEDLINE, Embase, Web of Science Core Collection, Cochrane Central Register of Controlled Trials, CINAHL, and Google Scholar) up to December 5, 2022.

**STUDY SELECTION** Studies were included in this systematic review and meta-analysis if they had a prospective cohort, retrospective cohort, or cross-sectional design and reported on the comparison between POC device(s) and LBB quantification in neonates aged 0 to 28 days. Point-of-care devices needed the following characteristics: portable, handheld, and able to provide a result within 30 minutes. This study was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-analyses reporting guideline.

**DATA EXTRACTION AND SYNTHESIS** Data extraction was performed by 2 independent reviewers into a prespecified, customized form. Risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Meta-analysis was performed of multiple Bland-Altman studies using the Tipton and Shuster method for the main outcome.

**MAIN OUTCOMES AND MEASURES** The main outcome was mean difference and limits of agreement in bilirubin levels between POC device and LBB quantification. Secondary outcomes were (1) turnaround time (TAT), (2) blood volumes, and (3) percentage of failed quantifications.

**RESULTS** Ten studies met the inclusion criteria (9 cross-sectional studies and 1 prospective cohort study), representing 3122 neonates. Three studies were considered to have a high risk of bias. The Bilistick was evaluated as the index test in 8 studies and the BiliSpec in 2. A total of 3122 paired measurements showed a pooled mean difference in total bilirubin levels of  $-14 \mu\text{mol/L}$ , with pooled 95% CIs of  $-106$  to  $78 \mu\text{mol/L}$ . For the Bilistick, the pooled mean difference was  $-17 \mu\text{mol/L}$  (95% CIs,  $-114$  to  $80 \mu\text{mol/L}$ ). Point-of-care devices were faster in returning results compared with LBB quantification, whereas blood volume needed was less. The Bilistick was more likely to have a failed quantification compared with LBB.

**CONCLUSIONS AND RELEVANCE** Despite the advantages that handheld POC devices offer, these findings suggest that the imprecision for measurement of neonatal bilirubin needs improvement to tailor neonatal jaundice management.

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Lauren E. H. Westenberg, MD, Division of Neonatology, Department of Pediatrics, Erasmus MC Sophia Children's Hospital, University Medical Center Rotterdam, Rotterdam, the Netherlands (l.westenberg@erasmusmc.nl).

JAMA Pediatr. 2023;177(5):479-488. doi:10.1001/jamapediatrics.2023.0059  
Published online March 13, 2023.

Neonatal jaundice is a condition caused by elevated unconjugated bilirubin levels. It occurs in up to 80% of neonates and is usually considered benign.<sup>1</sup> However, the transient imbalance between bilirubin production and clearance may unpredictably lead to a fast rise of unconjugated bilirubin to levels that are dangerous for the neonate's developing brain. High unconjugated bilirubin levels carry a risk for both acute and chronic permanent brain damage.<sup>2,3</sup> Although largely preventable, severe hyperbilirubinemia remains a major burden for neonatal health, especially in low-resource settings.<sup>1,4,5</sup>

Early and rapid diagnosis of hyperbilirubinemia is essential to prevent its deleterious effects. Neonatal hyperbilirubinemia can easily be treated by (intensive) phototherapy and, in severe cases, by exchange transfusion.<sup>6</sup> In many settings, identification of neonatal hyperbilirubinemia depends on screening via visual inspection of the neonate's skin color followed by selective laboratory-based bilirubin (LBB) quantification in serum, plasma, or whole blood.<sup>7,8</sup>

Laboratory-based bilirubin quantification can be performed using a variety of in vitro diagnostic instruments<sup>9,10</sup> and is the routine standard for diagnosing hyperbilirubinemia and indicating and monitoring treatment.<sup>11</sup> Yet, this approach can require up to 1500  $\mu$ L of blood and a fully equipped laboratory.<sup>12</sup> The turnaround time (TAT), ie, the time between deciding to quantify bilirubin and obtaining the result, can be rather long for LBB quantification. A TAT of hours is common in practice, especially in infants with jaundice cared for outside the hospital setting, in whom bilirubin levels may rise unnoticed.<sup>13</sup>

In low-resource settings, laboratories may be remote, poorly equipped, and not always able to provide an accurate LBB level.<sup>14-16</sup> As such, the diagnosis of jaundice in many neonates relies mainly on visual inspection, which is known to be unreliable.<sup>17,18</sup> Transcutaneous bilirubin is a fast and reliable method to estimate bilirubin levels in neonates and is useful as a screening instrument.<sup>19,20</sup> However, to establish the diagnosis of hyperbilirubinemia and determine the need for treatment, the standard remains a bilirubin quantification in blood. Therefore, the introduction of a reliable, quick, and affordable point-of-care (POC) bilirubin device that is easily transported is much needed, especially in low-resource settings or remote rural areas.<sup>16,21</sup>

Point-of-care testing is performed near or at the site of a patient and is increasingly being used in hospital and primary care settings. Point-of-care devices rapidly provide results and should facilitate earlier treatment initiation.<sup>22-24</sup> Moreover, POC devices require a smaller volume of whole blood (eg, 25-50  $\mu$ L). These characteristics make POC devices an appealing alternative to LBB provided that their reliability sufficiently approaches that of LBB quantification.<sup>4</sup>

We conducted a systematic review of studies that assessed the diagnostic accuracy of portable, handheld POC bilirubin devices. We performed a meta-analysis to provide an overall assessment of the diagnostic properties of POC bilirubin tests vs LBB quantification. Findings from this review could be instrumental in informing health care professionals about the usefulness and validation of POC devices for bilirubin quantification in everyday neonatal care.

## Key Points

**Question** What is the diagnostic accuracy of handheld point-of-care (POC) devices vs laboratory-based quantification for bilirubin in neonates?

**Findings** In this systematic review and meta-analysis of data from 10 studies representing 3122 neonates, POC devices tended to underestimate bilirubin levels in neonates, with a pooled mean difference of  $-14 \mu\text{mol/L}$ . Precision was limited compared with laboratory-based quantification, with pooled outer confidence bounds of  $-106$  to  $78 \mu\text{mol/L}$ .

**Meaning** Although handheld POC bilirubin devices allow for fast bilirubin measurements, this study's findings suggest that their imprecision limits widespread use for neonatal jaundice management, especially when accurate laboratory-based bilirubin quantification is available.

## Methods

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses reporting guideline.<sup>25,26</sup> Our review was registered in the International Prospective Register of Systematic Reviews (CRD42021289420).<sup>27</sup>

### Eligibility Criteria

Studies were eligible for inclusion if they were of a prospective cohort, retrospective cohort, or cross-sectional design and compared 1 or more POC devices and LBB measurement for quantifying neonatal bilirubin levels. Point-of-care bilirubin devices were eligible if they were portable, handheld, provided a test result within 30 minutes, and required whole blood or serum to quantify the total bilirubin level. Laboratory-based bilirubin quantification was considered the reference test. The index test needed to be conducted in the same time frame as the reference test. Studies evaluating newborns aged 0 to 28 days were considered eligible.

### Search Strategy and Selection Process

We searched 6 online data sources (Ovid MEDLINE, Embase, Web of Science Core Collection, Google Scholar, Cochrane Central Register of Controlled Trials, and CINAHL) up to December 5, 2022. The search strategy was designed in close collaboration with an experienced biomedical information specialist (W.M.B.) and specifically tailored for each database. The search strategy is available in the eMethods in Supplement 1.

The reference lists of relevant studies and their citations (through Google Scholar) were screened for additional potentially eligible studies that may have been previously missed. There were no restrictions imposed on language, publication date, or time frame of the study.

Records identified were imported into an EndNote (Clarivate) library through electronic extraction. After deduplication using the Bramer method,<sup>28</sup> 2 reviewers (L.E.H.W. and C.V.H.) manually identified any remaining duplicates. Subsequently, the 2 reviewers independently screened titles and

abstracts. Full-text reports of potentially eligible studies were retrieved and assessed for eligibility. Any disagreement was addressed and resolved through discussion or via consulting a third reviewer (J.V.B.).

### Data Extraction and Quality Assessment

Data extraction was performed by 2 independent reviewers (L.E.H.W. and C.V.H.) using a prespecified, customized form (eTable 1 in Supplement 1). Disagreements were resolved through discussion or via consulting a third reviewer (J.V.B.). In case of relevant missing information, the corresponding author of the respective study was contacted.

We used the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool<sup>29</sup> to evaluate risk of bias and applicability (eTable 2 in Supplement 1). The QUADAS-2 assessment was conducted by the 2 reviewers (L.E.H.W. and C.V.H.) independently. Consensus was reached for all studies on all domains.

### Outcomes

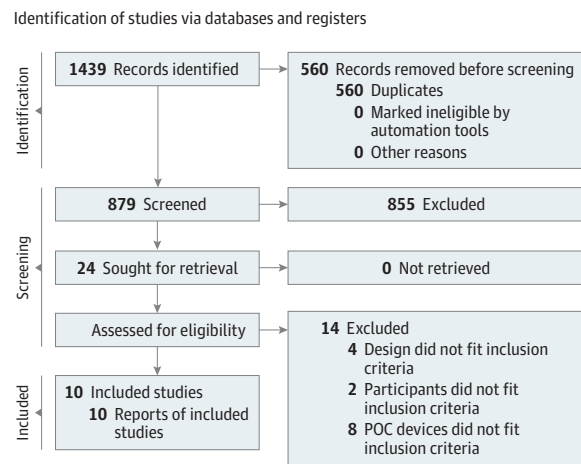
Our primary outcome of interest was the mean difference (ie, bias) and limits of agreement (LOAs) in total bilirubin level between the POC device (index test) and LBB quantification (reference test). Secondary outcomes were (1) TAT, (2) blood volumes, and (3) percentage of failed quantifications.

### Data Analysis

Our primary analyses involved pooling the overall Bland-Altman statistics to determine the diagnostic accuracy of POC device vs LBB quantification for assessing neonatal bilirubin. Our approach to performing the meta-analysis is based on the method of Tipton and Shuster<sup>30</sup> for meta-analysis of Bland-Altman studies, albeit incorporated in a bayesian framework. The advantage of a bayesian framework is making the modeling of several correlated studies easier. Furthermore, it relaxes the reliance on large-sample asymptotic theory, which is important in meta-analysis where the number of studies is often small. All priors were chosen to be weakly informative in order not to have too big an impact on the results.

Tipton and Shuster<sup>30</sup> provided a way to estimate population bias (expected difference between the 2 methods), construct pooled LOAs, and calculate the associated uncertainty. To summarize, the pooled LOAs are a function of 3 parameters: (1) population bias (the average difference between 2 tests in the population), (2) the average within-study variation, and (3) the variation in bias across studies. The study-specific estimates of the bias are pooled to give estimates of the population bias and the between-study variance of this bias. The average between-study variation is estimated by pooling the study-specific estimates. The population bias within-study variance and between-study variance can now be used to calculate the LOAs. By the usual definition, the LOAs are defined as the expected bias plus or minus 1.96 times the expected SD of the bias across individuals and study settings. Asymptotically, when the LOAs are constructed this way, the difference in outcomes between the methods when applied in a randomly selected new case has a 95% probability of being between the bounds. However, as

**Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 Flow Diagram, Including Searches of Databases, Registers, and Other Sources**



POC indicates point-of-care.

the usual LOA definition uses estimates in place of the true parameters of the population, in smaller samples (both in terms of studies and patients within studies), this is not the case. Therefore, we also calculated the outer confidence bounds (CBs) that fully take into account all uncertainty.

Besides pooling all studies, we also anticipated conducting subgroup analyses for each type of POC device separately. Furthermore, some included studies provided multiple Bland-Altman plots for different populations. To account for dependency of these observations, we added an extra hierarchical level to the model (so that there are both study-specific and population-within-study-specific effects of bias and variance).

The data were analyzed using R, version 4.1 (R Foundation for Statistical Computing) and STAN, version 2.26 (STAN Development Team) software. If needed, we converted units of bilirubin levels from mg/dL to  $\mu\text{mol/L}$  using a conversion factor of 17.1.

## Results

We identified 1439 records through database searching. After removing duplicates, titles and abstracts were screened for 879 records. Twenty-four records were selected for full-text assessment, of which 10 were included in qualitative and quantitative analyses<sup>12,31-39</sup> (Figure 1).

### Study Characteristics

Table 1 lists the study characteristics of the 10 included studies (representing 3122 neonates), of which 9 used a cross-sectional design<sup>12,31-38</sup> and 1 a prospective cohort design<sup>39</sup>; guidelines for phototherapy<sup>40-42</sup> used by some studies are also presented. Five studies were performed in Southeast Asia,<sup>12,31-34</sup> 3 in Africa,<sup>35-37</sup> 1 in Europe and Africa,<sup>38</sup> and 1 in

Table 1. Study Characteristics of Included Studies

Source	Country	Paired measurements	Study design	Study population		
				Inclusion criteria	Exclusion criteria	Clinical setting
Boo et al, <sup>12</sup> 2019	Malaysia	561	Cross-sectional	Gestational age >36 wk; clinical jaundice	With illness (not specified)	Neonatal nurseries of 2 government hospitals
Coda Zabetta et al, <sup>38</sup> 2013	Egypt	87	Cross-sectional	Term and near term; clinically requiring bilirubin determination	NR	Neonatal nurseries of 2 hospitals
	Italy	31				
Greco et al, <sup>36</sup> 2017	Egypt	126	Cross-sectional	Gestational age >36 wk; clinical jaundice	NR	Neonatal intensive care unit of a tertiary care referral center
Greco et al, <sup>39</sup> 2018	Egypt	130	Prospective cohort	Gestational age ≥35 wk; postnatal age <28 d; visually jaundiced or signs of acute bilirubin encephalopathy routine screening of bilirubin	NR	17 Medical centers
	Indonesia	530				
	Nigeria	168				
	Vietnam	630				
Kamineni et al, <sup>31</sup> 2020	India	198	Cross-sectional	Preterm and term neonates; clinically requiring bilirubin measurement; at risk for jaundice	Prior exchange or blood transfusion	Tertiary hospital for women and newborns
Keahey et al, <sup>35</sup> 2017	Malawi	63	Cross-sectional	Age <28 d; at risk for jaundice	NR	Government hospital
Rohsiswatmo et al, <sup>34</sup> 2018	Indonesia	94	Cross-sectional	Gestational age <35 wk; postnatal age ≤14 wk; indication for phototherapy based on IDAI guidelines <sup>40</sup>	After phototherapy or exchange transfusion	Neonatal nursery of government hospital
Sampurna et al, <sup>32</sup> 2021	Indonesia	126	Cross-sectional	Gestational age ≥32 wk; postnatal age ≤14 d; birth weight ≥1500 g; clinical jaundice as per Kramer scale (any Kramer score >0) <sup>41</sup>	Phototherapy in the preceding 24 h; respiratory or circulatory insufficiency; severe congenital abnormalities	Government hospital
Shapiro et al, <sup>37</sup> 2022	Malawi	326	Cross-sectional	Hospitalized; at risk for jaundice	NR	Neonatal nurseries of 2 government hospitals
Thielemans et al, <sup>33</sup> 2018	Thailand	52	Cross-sectional	Gestational age ≥35 wk; clinical jaundice as per Kramer scale (score >3) <sup>41</sup> ; previous borderline serum bilirubin measurement (ie, ≤50 μmol/L below the treatment threshold of NICE guidelines <sup>42</sup> )	NR	Field clinic

Abbreviations: IDAI, Ikatan Dokter Anak Indonesia (Indonesian Pediatric Society); NICE, National Institute for Health and Care Excellence; NR, not reported.

Africa and Southeast Asia.<sup>39</sup> Studies were published between 2012 and 2022. Characteristics of the POC devices and LBB quantification are available in eTable 3 in Supplement 1.

The studies reported a total of 3122 paired measurements. Of the 10 included studies, 8 studied the Bilistick (Bilimetrix srl), a POC test for total bilirubin quantification in 25 μL of whole blood.<sup>12,31-34,36,38,39</sup> The Bilistick consists of a handheld reflectance reader and test strips, which are composed of a blood-plasma filter and nitrocellulose membrane. The blood is loaded on the test strip and filtered. Next, the plasma diffuses to the membrane where bilirubin is measured by the reader through reflectance spectroscopy. The Bilistick requires calibration every 6 months. Two studies evaluated the performance of BiliSpec (3rd Stone Design), a POC test for total bilirubin quantification in 50 μL of whole blood.<sup>35,37</sup> BiliSpec, now called BiliDx, consists of a handheld reader and lateral flow card. The lateral flow card separates the blood and plasma (using a fiber plasma separation membrane) and stabilizes the sample so that the bilirubin concentration remains constant over time. Subsequently, the

reader measures the light transmitted through the plasma on the card, and the bilirubin level is estimated. BiliSpec calibration cards are housed inside the reader. The calibration process was performed daily during the trials.<sup>35,37</sup> There was substantial variation in the reference methods used, and these included blood gas and chemistry analyzers.

### Quality Assessment

Four studies<sup>31,32,34,36</sup> were considered low risk of poor quality in all domains (Figure 2). Flow and timing were considered at high risk of bias in 3 studies where different laboratory reference methods were used within the study population, potentially signaling verification bias.<sup>12,38,39</sup>

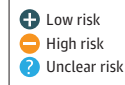
### Diagnostic Accuracy of POC Devices Compared With LBB Quantification

All 8 studies involving the Bilistick reported an underestimation of the POC device compared with LBB quantification for assessing neonatal bilirubin.<sup>12,31-34,36,38,39</sup> The 2 studies evalu-



Figure 2. Quality Assessment of Diagnostic Accuracy Studies 2 Diagram

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Boo et al, <sup>12</sup> 2019	+	+	+	-	+	+	+
Coda Zabetta et al, <sup>38</sup> 2013	+	+	+	-	+	+	+
Greco et al, <sup>36</sup> 2017	+	+	+	+	+	+	+
Greco et al, <sup>39</sup> 2018	+	+	+	-	+	+	+
Kamineni et al, <sup>31</sup> 2020	+	+	+	+	+	+	+
Keahey et al, <sup>35</sup> 2017	?	+	?	+	?	+	?
Rohsiswatmo et al, <sup>34</sup> 2018	+	+	+	+	+	+	+
Sampurna et al, <sup>32</sup> 2021	+	+	+	+	+	+	+
Shapiro et al, <sup>37</sup> 2022	+	?	+	+	+	+	?
Thielemans et al, <sup>33</sup> 2018	+	+	?	+	+	+	?



ating the BiliSpec had contrasting outcomes with regard to bias<sup>35,37</sup> (Table 2). Results from 2 studies were presented in a single Bland-Altman plot, even though the data were retrieved from patients in different hospitals that used a different reference method to measure total serum bilirubin.<sup>12,38</sup> One study was performed in 17 medical centers in 4 countries, and results were presented in 4 Bland-Altman plots for these 4 countries.<sup>39</sup> Thirteen effect estimates derived from the 10 studies evaluating POC bilirubin devices were meta-analyzed. The pooled mean difference between the POC bilirubin device and LBB quantification was  $-14 \mu\text{mol/L}$ , with the pooled 95% LOA ranging from  $-86$  to  $57 \mu\text{mol/L}$  and 95% CBs of  $-106$  to  $78 \mu\text{mol/L}$  (Figure 3A). Meta-analysis was performed for the Bilistick separately. Eleven effect estimates from the 8 studies were evaluated for the Bilistick. The pooled estimate of the mean difference between the Bilistick and LBB quantification was  $-17 \mu\text{mol/L}$ , with the pooled 95% LOA ranging from  $-91$  to  $57 \mu\text{mol/L}$  and 95% CBs of  $-114$  to  $80 \mu\text{mol/L}$ . Due to the small number of eligible studies evaluating the BiliSpec ( $n = 2$ ), we could not conduct our pre-planned subgroup analysis. eTable 4 in Supplement 1 shows the summary of findings.

For the multilevel meta-analysis of these estimates, we clustered the results from Greco et al.<sup>39</sup> Results from the multilevel meta-analysis were similar to those from the primary analysis (Figure 3B): For all POC devices, the mean difference was  $-17 \mu\text{mol/L}$ , with the pooled LOA ranging from  $-89$  to  $55 \mu\text{mol/L}$  and 95% CBs of  $-119$  to  $87 \mu\text{mol/L}$ . For the Bilistick, the mean difference was  $-17 \mu\text{mol/L}$ , with the pooled 95% LOA ranging from  $-89$  to  $55 \mu\text{mol/L}$  and 95% CBs of  $-119$  to  $86 \mu\text{mol/L}$ .

### Turnaround Time

Only 1 study by Boo et al<sup>12</sup> reported on the TAT difference (Table 2). In the 2 participating hospitals, the median TATs of the LBB quantification were 98 minutes (range, 24-424 minutes) and 114 minutes (range, 34-1039 minutes). Compared with the mean TAT of the Bilistick of 2 minutes, the hospital LBB quantification was approximately 60 times longer ( $P < .001$ ). Eight studies reported on the time interval between when the blood was loaded on the test strip and when the bilirubin level was displayed on the POC's screen: 2

minutes for the BiliSpec<sup>35</sup> and between 100 seconds and 3 minutes for the Bilistick.<sup>12,31,33,34,36,38,39</sup>

### Blood Volumes

All studies evaluating the Bilistick reported a blood volume of  $25 \mu\text{L}$  for sampling<sup>12,31-34,36,38,39</sup>; blood volume for the BiliSpec was  $50 \mu\text{L}$ .<sup>35,37</sup> Four studies reported blood volumes ranging between 50 and  $1500 \mu\text{L}$  for the reference test in the laboratory<sup>12,33,34,37</sup> (eTable 3 in Supplement 1). Blood volume needed was 40 to 60 times less for POC devices than for LBB quantification.

### Percentage of Failed Quantifications

The reported percentage of failed quantifications of the Bilistick varied between 1.5% and 48.6% across studies, with a median of 8.75% (IQR, 5.02%-24.07%).<sup>12,31-33,36,39</sup> Common error messages were high hematocrit ( $>65\%$ ) and hemolysis. One study reported 9% failed measurements of the BiliSpec due to the strip not adequately being filled.<sup>37</sup> The reported percentage of error rates of the laboratory quantification varied between 1.3% and 2.5% and were due to technical malfunction.<sup>12,36</sup>

## Discussion

In this systematic review and meta-analysis, we assessed the accuracy of portable, handheld POC bilirubin devices, ie, the Bilistick and BiliSpec. Overall, the POC devices tended to underestimate the neonatal bilirubin level compared with conventional LBB quantification methods. Furthermore, the pooled estimates revealed that these POC bilirubin devices are imprecise, where the calculated outer confidence bounds were substantial, ie,  $-106$  to  $78 \mu\text{mol/L}$  for all of the POC devices and  $-114$  to  $80 \mu\text{mol/L}$  for the Bilistick. The multilevel meta-analysis showed similar results to the primary analysis, strengthening our confidence in the results. Bilirubin test results were up to approximately 60 times faster through POC devices compared with LBB quantification, whereas the volumes of the used blood samples were 40 to 60 times less using POC devices. Failed measurements occurred in up to 49% of measurements for the Bilistick,<sup>33</sup> which was attributed to

Table 2. Primary and Secondary Outcomes Reported in Included Studies

Source	MD (bias) (SD) <sup>a</sup>	Corresponding LOAs, μmol/L <sup>a</sup>		Variability of POC device across different bilirubin concentrations	TAT, POC vs LBB	Percentage of failed quantifications, POC vs LBB	Missed diagnosis by POC
		UL	LL				
Boo et al, <sup>12</sup> 2019	-26.5 (-29.4)	31.2	-84.1	Underestimation was also observed for values >300 μmol/L	LBB quantification: 98 (range, 24-424) min and 114 (range, 34-1039); POC: 2 min; hospital LBB quantification was significantly longer (P < .001)	Bilistick: 15.9 due to high hematocrit and blood clotting; Jendrassik-Grof method: 1.3; sample rejected by laboratory method	Above treatment threshold, the sensitivity was 0.74, ie, approximately 26% missed diagnoses; sensitivity lower at higher Bilistick cutoff values
Coda Zabetta et al, <sup>38</sup> 2013	-10.3 (24.1)	38.0	-58.7	Underestimation was observed more frequently at values >342 μmol/L	NR	Bilistick: NR; Jendrassik-Grof method and diazo method: NR	NR
Greco et al, <sup>36</sup> 2017	-22.2 (39.3)	56.4	-99.2	No specific comment	NR	Bilistick: 6.8 due to technical problems; diazo method: 2.5 due to technical problems	NR
Greco et al, <sup>39</sup> 2018	Egypt: -29.1 (49.6)	68.4	-126.5	Underestimation was stable over a wide range of bilirubin values	NR	Bilistick: 1.5 due to technical problems; diazo method or direct spectrophotometry: NR	Above treatment threshold, the sensitivity was 0.70, ie, approximately 30% missed diagnoses
	Indonesia: -13.7 (27.4)	39.3	-68.4				
	Nigeria: -13.7 (58.1)	102.6	-128.2				
	Vietnam: -15.4 (39.3)	63.3	-94.0				
Kamineni et al, <sup>31</sup> 2020	-8.6 (75.2)	140.2	-155.6	Underestimation was observed across all ranges of bilirubin levels	NR	Bilistick: 6.2 due to hemolysis; direct spectrophotometry: NR	NR
Keahey et al, <sup>35</sup> 2017	5.1 (17.1)	37.6	-29.1	No specific comment	NR	BiliSpec: NR; direct spectrophotometry: NR	NR
Rohsiswatmo et al, <sup>34</sup> 2018	-25.6 (25.1) <sup>b</sup>	23.6	-74.9	Underestimation was observed more frequently at higher bilirubin levels	NR	Bilistick: NR; chemical oxidation method: NR	If treatment decisions would have been based on the POC device, then 11 of 94 infants who needed treatment would have been missed (approximately 12%)
Sampurna et al, <sup>32</sup> 2021	-11.0 (46.0)	79.0	-101.0	No specific comment	NR	Bilistick: 10.7 due to different device error messages; diazo method: NR	If treatment decisions would have been based on the POC device, then 10 of 39 infants who needed treatment would have been missed (approximately 26%)
Shapiro et al, <sup>37</sup> 2022	-8.2 (NR)	70.4	-87.0	No specific comment	NR	BiliSpec: 9 due to strip not adequately filling; direct spectrophotometry: 9 due to different device error messages or missing measurements	If treatment decisions would have been based on the POC device, 90.7% would have resulted in the same decision as the LBB quantification, ie, approximately 9% missed diagnoses
Thielemans et al, <sup>33</sup> 2018	-20.0 (NR)	18.0	-59.0	No specific comment	NR	Bilistick: 48.6% due to different device error messages; direct spectrophotometry: NR	NR

Abbreviations: LBB, laboratory-based bilirubin; LL, lower limit; LOA, limit of agreement; MD (bias), mean difference; NR, not reported; POC, point-of-care; TAT, turnaround time; UL, upper limit.

<sup>a</sup> The MD (bias) and corresponding LOAs of Bilistick bilirubin - LBB, meaning

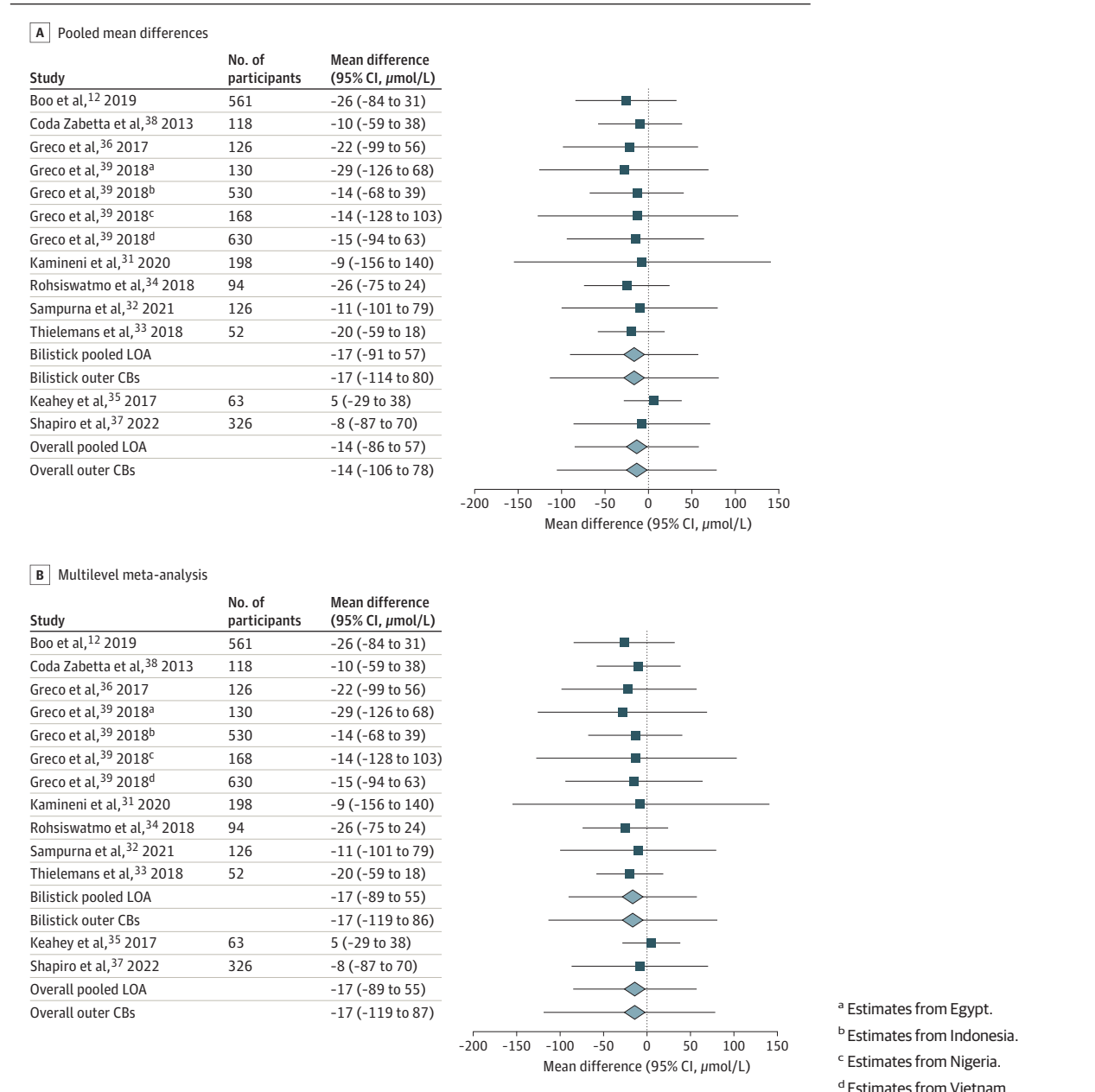
that the difference between Bilistick bilirubin and LBB is plotted against the mean of Bilistick bilirubin and LBB. We converted units of bilirubin levels from mg/dL to μmol/L using a conversion factor of 17.1 (ie, 1 mg/dL = 17.1 μmol/L).

<sup>b</sup> Quantification before phototherapy treatment.

weather conditions (high humidity) and seemed not representative for updated versions in similar climates.<sup>32</sup> In studies comparing error rates between POC device and LBB quantification, the POC devices were more likely to fail.

The inaccuracy of the POC device quantification should be considered in light of the accuracy of conventional LBB quantification methods that are generally considered as the clinical reference methods for bilirubin quantification.

**Figure 3. Forest Plots of Pooled Mean Difference (Bias), Limits of Agreement (LOAs), and Outer Confidence Bounds (CBs)**



Although viewed as the reference standard, several studies reported rather high interlaboratory variability and inaccuracy of LBB levels measured in value-assigned commutable specimens as well as in patient samples, signaling inaccuracy and imprecision.<sup>10,43-47</sup> The substantial inconsistencies encountered when quantifying bilirubin on different commonly used multianalysis instruments underline the timelessness of Mather’s statement that “bilirubin determinations are perhaps the most notoriously unreliable of any in clinical chemistry.”<sup>48(p350)</sup> The inaccuracy and high variability of POC bilirubin test results may, at least in part, originate from hemolysis or be affected by the relatively high hemoglobin concentration in the newborn.<sup>33</sup> To this end,

the Bilistick displays a warning message in case of too much hemolysis.

The question remains whether the POC bilirubin devices are fit for clinical decision making. Despite the short TAT, ease of use, and relatively low costs, their use should be conditional upon the reliability to produce accurate bilirubin results. Theoretically, use of POC bilirubin devices entails a risk of missing any neonate with jaundice who actually needs phototherapy or, in the case of overestimation, starting phototherapy too early. The risk of misclassification was encountered and acknowledged by several of the studies.<sup>32,34,35,37,39</sup> Moreover, POC devices require near-perfect conditions for optimal use and results in terms of humidity, preanalytic con-



ditions (eg, saturation of test strip or membrane), and hematocrit. Despite the risk of misclassification, POC bilirubin devices might be useful when diagnosing neonatal hyperbilirubinemia solely on visual inspection or when laboratory facilities are remote, which might be the case in some low-resource settings.

Precise quantification of POC bilirubin is highly clinically relevant because treatment thresholds in current international guidelines for the management of neonatal hyperbilirubinemia are based on bilirubin concentrations.<sup>49</sup> To avoid unnecessary escalation of care, it is important that the measured bilirubin level approaches the true level, irrespective of the applied method. Our recommendations are based on 10 studies, of which 3 were considered as having a high risk of bias based on QUADAS-2 assessment (Figure 2). Future research should focus on assessing the proportional bias of the POC bilirubin devices for multiple concentrations of bilirubin. Our findings suggest that handheld POC bilirubin devices still need to be optimized in order to overcome their limitations. To support their future assessment, the total allowable error and clinically relevant limits of bilirubin quantification should be defined in international guidelines, including specific recommendations for handheld POC bilirubin devices. To harmonize POC bilirubin results with LBB methods, incorporation of POC bilirubin devices into an external quality assessment program seems sensible analogous to existing external quality assessment for LBB methods. While awaiting such data, test results of POC bilirubin readings should be interpreted cautiously.

### Strengths and Limitations

To our knowledge, this systematic review and meta-analysis is the first of existing data on 2 POC devices for neonatal bilirubin quantification, which are currently available and CE (Conformité Européenne) marked. We used a comprehensive search strategy, including several electronic databases and reference and citation searches. Moreover, we used a unique approach for the data synthesis wherein correlation coefficients were used to measure the relationship between variables without giving any information on the meaningful clinical association.<sup>50</sup> In 1983, Bland and Altman<sup>51</sup> underlined these shortcomings and provided a new approach to method comparison studies. We focused on Bland-Altman method comparison studies, which are frequently used in laboratory medicine for the validation of a new method.<sup>52</sup> This systematic review is one of the first meta-analyses to use the unified framework for Bland-Altman meta-analysis provided by Tip-

ton and Shuster,<sup>30</sup> allowing us to calculate the LOAs and their associated outer confidence bounds (measures of uncertainty). Our data provide clinically relevant information for all contemporary and future users of POC bilirubin instruments involved in the care of infants with jaundice.

We acknowledge several limitations of this study. First, we were not able to calculate the variability across different bilirubin concentrations of the POC measurements, as the necessary data were not available from the original reports included in our review. Two studies reported that underestimation by the POC device was observed more frequently at higher bilirubin levels<sup>34,38</sup> (eTable 3 in Supplement 1). This underestimation by the device is important because it carries a risk of undertreatment. Especially at high levels of bilirubin, POC bilirubin quantification should be accurate to minimize the risk of undertreatment. Second, overall data on the diagnostic accuracy to indicate the need for treatment could not be retrieved. The largest study on the Bilistick reported a sensitivity of 70.8% (95% CI, 65.0%-75.7%) for diagnosis of hyperbilirubinemia requiring treatment, with a specificity of 98.5% (95% CI, 97.7%-99.1%) and similar positive and negative predictive values of approximately 93%,<sup>39</sup> and another multicenter study showed sensitivity, specificity, and positive and negative predictive values of 0.74, 0.84, 0.67, and 0.88, respectively, for any bilirubin value above treatment thresholds.<sup>12</sup> Five studies reported on the number of infants with hyperbilirubinemia necessitating treatment that would have been missed by the POC device (9%-30%)<sup>12,32,34,37,39</sup> (Table 2). Third, the variety of the applied clinical reference method with reported inaccuracies for bilirubin quantification may be considered as a limitation of this systematic review, although it may also strengthen the external validation of our findings. Fourth, most of the 3122 neonates represented in this review were late preterm and term neonates. As such, generalizability of the findings to preterm and very preterm newborns is limited.

### Conclusions

Quantification of bilirubin levels using POC devices requires less blood volume and produces faster results compared with conventional LBB quantification. Despite these advantages, our findings in this systematic review and meta-analysis suggest that the percentage of failed quantifications and their imprecision for measurement of neonatal bilirubin need to be improved in POC devices.

#### ARTICLE INFORMATION

**Accepted for Publication:** January 11, 2023.

**Published Online:** March 13, 2023.  
doi:10.1001/jamapediatrics.2023.0059

**Author Affiliations:** Division of Neonatology, Department of Pediatrics, Erasmus MC Sophia Children's Hospital, University Medical Center Rotterdam, Rotterdam, the Netherlands (Westenberg, Been, Reiss); Department of Obstetrics and Gynecology, Erasmus MC Sophia Children's Hospital, University Medical Center

Rotterdam, Rotterdam, the Netherlands (Been, Willemsen, Steegers); Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands (Been); Department of Biostatistics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands (Willemsen); Department of Clinical Chemistry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands (Vis, Tintu); Medical Library, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands (Bramer); Division of Neonatology, Department of Pediatrics,

Beatrix Children's Hospital, University Medical Center Groningen, Groningen, the Netherlands (Dijk, Hulzebos).

**Author Contributions:** Drs Westenberg and Hulzebos had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Concept and design:** Westenberg, Been, Tintu, Steegers, Reiss, Hulzebos.

**Acquisition, analysis, or interpretation of data:** Westenberg, Been, Willemsen, Vis, Tintu, Bramer,

Dijk, Reiss, Hulzebos.

*Drafting of the manuscript:* Westenberg, Been, Reiss, Hulzebos.

*Critical revision of the manuscript for important intellectual content:* Been, Willemsen, Vis, Tintu, Bramer, Dijk, Steegers, Reiss, Hulzebos.

*Statistical analysis:* Westenberg, Willemsen.

*Obtained funding:* Been.

*Administrative, technical, or material support:* Tintu, Bramer, Hulzebos.

*Supervision:* Been, Willemsen, Vis, Tintu, Dijk, Steegers, Reiss, Hulzebos.

**Conflict of Interest Disclosures:** None reported.

**Funding/Support:** This work was supported by grant 543002011 from The Netherlands Organization for Health Research and Development (Dr Been).

**Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See [Supplement 2](#).

**Additional Contributions:** The authors thank Elizabeth Tipton, PhD (Northwestern University), and Jonathan Shuster, PhD (University of Florida), for help with the data synthesis. Drs Tipton and Shuster were not compensated for their contributions.

## REFERENCES

- Bhutani VK, Zipursky A, Blencowe H, et al. Neonatal hyperbilirubinemia and Rhesus disease of the newborn: incidence and impairment estimates for 2010 at regional and global levels. *Pediatr Res*. 2013;74(suppl 1):86-100. doi:10.1038/pr.2013.208
- Le Pichon JB, Riordan SM, Watchko J, Shapiro SM. The neurological sequelae of neonatal hyperbilirubinemia: definitions, diagnosis and treatment of the kernicterus spectrum disorders (KSDs). *Curr Pediatr Rev*. 2017;13(3):199-209. doi:10.2174/1573396313666170815100214
- Johnson L, Bhutani VK. The clinical syndrome of bilirubin-induced neurologic dysfunction. *Semin Perinatol*. 2011;35(3):101-113. doi:10.1053/j.semperi.2011.02.003
- Greco C, Arnolda G, Boo NY, et al. Neonatal jaundice in low- and middle-income countries: lessons and future directions from the 2015 Don Ostrow Trieste Yellow Retreat. *Neonatology*. 2016; 110(3):172-180. doi:10.1159/000445708
- Olusanya BO, Ogunlesi TA, Slusher TM. Why is kernicterus still a major cause of death and disability in low-income and middle-income countries? *Arch Dis Child*. 2014;99(12):1117-1121. doi:10.1136/archdischild-2013-305506
- Muchowski KE. Evaluation and treatment of neonatal hyperbilirubinemia. *Am Fam Physician*. 2014;89(11):873-878.
- de Boer J, Zondag L. *Multidisciplinaire richtlijn Postnatale Zorg: Verloskundige basiszorg voor moeder en kind*. Koninklijke Nederlandse Organisatie van Verloskundigen; 2018.
- Dijk PH, de Vries TW, de Beer JJA; Dutch Pediatric Association. Richtlijn 'Preventie, diagnostiek en behandeling van hyperbilirubinemie bij de pasgeborene, geboren na een zwangerschapsduur van meer dan 35 weken'. *Ned Tijdschr Geneesk*. 2009;153:A93.
- Hulzebos CV, Vitek L, Coda Zabetta CD, et al. Diagnostic methods for neonatal hyperbilirubinemia: benefits, limitations, requirements, and novel developments. *Pediatr Res*. 2021;90(2): 277-283. doi:10.1038/s41390-021-01546-y
- Greene DN, Liang J, Holmes DT, Resch A, Lorey TS. Neonatal total bilirubin measurements: still room for harmonization. *Clin Biochem*. 2014;47(12): 1112-1115. doi:10.1016/j.clinbiochem.2014.04.001
- Olusanya BO, Kaplan M, Hansen TWR. Neonatal hyperbilirubinemia: a global perspective. *Lancet Child Adolesc Health*. 2018;2(8):610-620. doi:10.1016/S2352-4642(18)30139-1
- Boo NY, Chang YF, Leong YX, et al. The point-of-care Bilistick method has very short turn-around-time and high accuracy at lower cutoff levels to predict laboratory-measured TSB. *Pediatr Res*. 2019;86(2):216-220. doi:10.1038/s41390-019-0304-0
- van der Geest BAM, Rosman AN, Bergman KA, et al. Severe neonatal hyperbilirubinemia: lessons learnt from a national perinatal audit. *Arch Dis Child Fetal Neonatal Ed*. 2022;107(5):527-532. doi:10.1136/archdischild-2021-322891
- Olusanya BO, Ogunlesi TA, Kumar P, et al. Management of late-preterm and term infants with hyperbilirubinemia in resource-constrained settings. *BMC Pediatr*. 2015;15:39. doi:10.1186/s12887-015-0358-z
- Anticona Huaynate CF, Pajuelo Travezaño MJ, Correa M, et al. Diagnostics barriers and innovations in rural areas: insights from junior medical doctors on the frontlines of rural care in Peru. *BMC Health Serv Res*. 2015;15:454. doi:10.1186/s12913-015-1114-7
- Florkowski C, Don-Wauchope A, Gimenez N, Rodriguez-Capote K, Wils J, Zemlin A. Point-of-care testing (POCT) and evidence-based laboratory medicine (EBLM)—does it leverage any advantage in clinical decision making? *Crit Rev Clin Lab Sci*. 2017;54(7-8):471-494. doi:10.1080/10408363.2017.1399336
- Moyer VA, Ahn C, Sneed S. Accuracy of clinical judgment in neonatal jaundice. *Arch Pediatr Adolesc Med*. 2000;154(4):391-394. doi:10.1001/archpedi.154.4.391
- Okolie F, South-Paul JE, Watchko JF. Combating the hidden health disparity of kernicterus in Black infants: a review. *JAMA Pediatr*. 2020;174(12): 1199-1205. doi:10.1001/jamapediatrics.2020.1767
- Dani C, Hulzebos CV, Tiribelli C. Transcutaneous bilirubin measurements: useful, but also reproducible? *Pediatr Res*. 2021;89(4):725-726. doi:10.1038/s41390-020-01242-3
- Okwundu C, Bhutani VK, Smith J, Esterhuizen TM, Wiysonge C. PredischARGE transcutaneous bilirubin screening reduces readmission rate for hyperbilirubinemia in diverse South African newborns: a randomised controlled trial. *S Afr Med J*. 2020;110(3):249-254. doi:10.7196/SAMJ.2020.v110i3.14186
- Slusher TM, Zipursky A, Bhutani VK. A global need for affordable neonatal jaundice technologies. *Semin Perinatol*. 2011;35(3):185-191. doi:10.1053/j.semperi.2011.02.014
- Abel G. Current status and future prospects of point-of-care testing around the globe. *Expert Rev Mol Diagn*. 2015;15(7):853-855. doi:10.1586/14737159.2015.1060126
- McPartlin DA, O'Kennedy RJ. Point-of-care diagnostics, a major opportunity for change in traditional diagnostic approaches: potential and limitations. *Expert Rev Mol Diagn*. 2014;14(8): 979-998. doi:10.1586/14737159.2014.960516
- Patel K, Suh-Lailam BB. Implementation of point-of-care testing in a pediatric healthcare setting. *Crit Rev Clin Lab Sci*. 2019;56(4):239-246. doi:10.1080/10408363.2019.1590306
- Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372(71):n71. doi:10.1136/bmj.n71
- Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Syst Rev*. 2012;1:7. doi:10.1186/2046-4053-1-7
- Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc*. 2016;104(3):240-243. doi:10.3163/1536-5050.104.3.014
- Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009
- Tipton E, Shuster J. A framework for the meta-analysis of Bland-Altman studies based on a limits of agreement approach. *Stat Med*. 2017;36(23):3621-3635. doi:10.1002/sim.7352
- Kamineni B, Tanniru A, Vardhelli V, et al. Accuracy of Bilistick (a point-of-care device) to detect neonatal hyperbilirubinemia. *J Trop Pediatr*. 2020;66(6):630-636. doi:10.1093/tropej/fmaa026
- Sampurna MTA, Rani SAD, Sauer PJJ, Bos AF, Dijk PH, Hulzebos CV. Diagnostic properties of a portable point-of-care method to measure bilirubin and a transcutaneous bilirubinometer. *Neonatology*. 2021;118(6):678-684. doi:10.1159/000518653
- Thielemans L, Hashmi A, Priscilla DD, et al. Laboratory validation and field usability assessment of a point-of-care test for serum bilirubin levels in neonates in a tropical setting. *Wellcome Open Res*. 2018;3:110. doi:10.12688/wellcomeopenres.14767.1
- Rohsiswatmo R, Oswari H, Amantito R, et al. Agreement test of transcutaneous bilirubin and Bilistick with serum bilirubin in preterm infants receiving phototherapy. *BMC Pediatr*. 2018;18(1):315. doi:10.1186/s12887-018-1290-9
- Keahey PA, Simeral ML, Schroder KJ, et al. Point-of-care device to diagnose and monitor neonatal jaundice in low-resource settings. *Proc Natl Acad Sci U S A*. 2017;114(51):E10965-E10971. doi:10.1073/pnas.1714020114
- Greco C, Iskander IF, Akmal DM, et al. Comparison between Bilistick system and

transcutaneous bilirubin in assessing total bilirubin serum concentration in jaundiced newborns.

*J Perinatol*. 2017;37(9):1028-1031. doi:10.1038/jp.2017.94

37. Shapiro A, Anderson J, Mtenthaonga P, et al. Evaluation of a point-of-care test for bilirubin in Malawi. *Pediatrics*. 2022;150(2):e2021053928. doi:10.1542/peds.2021-053928

38. Coda Zabetta CD, Iskander IF, Greco C, et al. Bilistick: a low-cost point-of-care system to measure total plasma bilirubin. *Neonatology*. 2013;103(3):177-181. doi:10.1159/000345425

39. Greco C, Iskander IF, El Houchi SZ, et al; Study Team. Diagnostic performance analysis of the point-of-care Bilistick system in identifying severe neonatal hyperbilirubinemia by a multi-country approach. *EClinicalMedicine*. 2018;1:14-20. doi:10.1016/j.eclinm.2018.06.003

40. Kosim MS, Yunanto A, Dewi R, Sarosa GI, Usman A. *Buku Ajar Neonatologi*. Ikatan Dokter Anak Indonesia; 2008.

41. Kramer LI. Advancement of dermal icterus in the jaundiced newborn. *AJDC*. 1969;118(3):454-458. doi:10.1001/archpedi.1969.02100040456007

42. National Collaborating Centre for Women's and Children's Health. *Neonatal Jaundice: Clinical Guideline*. National Institute for Health and Care Excellence; 2010. Accessed January 8, 2022. <https://www.nice.org.uk/cg98/evidence/full-guideline-245411821>

43. Vreman HJ, Verter J, Oh W, et al. Interlaboratory variability of bilirubin measurements. *Clin Chem*. 1996;42(6, pt 1):869-873. doi:10.1093/clinchem/42.6.869

44. Cobbaert C, Weykamp C, Hulzebos CV. Bilirubin standardization in the Netherlands: alignment within and between manufacturers. *Clin Chem*. 2010;56(5):872-873. doi:10.1373/clinchem.2009.142059

45. Lo SF, Jendrzyczak B, Doumas BT; College of American Pathologists. Laboratory performance in neonatal bilirubin testing using commutable specimens: a progress report on a College of American Pathologists study. *Arch Pathol Lab Med*. 2008;132(11):1781-1785. doi:10.5858/132.11.1781

46. Grohmann K, Roser M, Rolinski B, et al. Bilirubin measurement for neonates: comparison of 9 frequently used methods. *Pediatrics*. 2006;117(4):1174-1183. doi:10.1542/peds.2005-0590

47. Mather A. Reliability of bilirubin determinations in icterus of the newborn infant. *Pediatrics*. 1960;26:350-354. doi:10.1542/peds.26.3.350

48. Hollis S. Analysis of method comparison studies. *Ann Clin Biochem*. 1996;33(pt 1):1-4. doi:10.1177/000456329603300101

49. Kemper AR, Newman TB, Slaughter JL, et al. Clinical Practice guideline revision: management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics*. 2022;150(3):e2022058859 doi:10.1542/peds.2022-058859

50. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310. doi:10.1016/S0140-6736(86)90837-8

51. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141-151. doi:10.11613/BM.2015.015

52. Maisels MJ, Bhutani VK, Bogen D, Newman TB, Stark AR, Watchko JF. Hyperbilirubinemia in the newborn infant > or =35 weeks' gestation: an update with clarifications. *Pediatrics*. 2009;124(4):1193-1198. doi:10.1542/peds.2009-0329