



## RESEARCH ARTICLE

# Laboratory validation and field usability assessment of a point-of-care test for serum bilirubin levels in neonates in a tropical setting [version 1; peer review: 2 approved, 1 approved with reservations]

Laurence Thielemans <sup>1,2</sup>, Ahmar Hashmi <sup>1</sup>, Dah Dah Priscilla<sup>1</sup>, Moo Kho Paw<sup>1</sup>, Tekel Pimolsorntong<sup>1</sup>, Thatsanun Ngerseng <sup>3</sup>, Bart Van Overmeire<sup>2</sup>, Stephane Proux <sup>1</sup>, François Nosten <sup>1,4</sup>, Rose McGready <sup>1,4</sup>, Verena I. Carrara <sup>1</sup>, Germana Bancone <sup>1,4</sup>

<sup>1</sup>Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, 63110, Thailand

<sup>2</sup>Neonatology, Cliniques Universitaires de Bruxelles - Hôpital Erasme, Université Libre de Bruxelles, Bruxelles, 1070, Belgium

<sup>3</sup>Mahidol-Oxford Tropical Medicine Research Unit (MORU), Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand

<sup>4</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK

**V1** First published: 04 Sep 2018, 3:110  
<https://doi.org/10.12688/wellcomeopenres.14767.1>  
Latest published: 23 Nov 2018, 3:110  
<https://doi.org/10.12688/wellcomeopenres.14767.2>

## Abstract





**Background:** Screening and monitoring serum bilirubin (SBR) in neonates is crucial to prevent neonatal hyperbilirubinemia (NH)-associated morbidity and mortality worldwide. A lack of resources is often a barrier for measuring SBR in developing countries. Reliable, cost-effective, easy to use point-of-care (POC) SBR tests are needed. This study aimed to evaluate the technical accuracy and usability of the Bilistick System (BS), a new bilirubin POC test, in a tropical setting.

**Methods:** This was a mixed-methods study, including laboratory validation of the BS, direct observation of technical procedures as performed by the midwives and midwives' assessment of the device's easiness of use through focus group discussions (FGD) and a self-administered questionnaire. The study was conducted in a field clinic of the Shoklo Malaria Research Unit along the Thailand-Myanmar border between January and December 2017.

**Results:** A total of 173 samples were tested at a median age of 4 days. BS generated an error message—providing no SBR readout—in 48.6% of the tests performed. For the tests that yielded a result, the correlation coefficient (95% CI) between BS and routine laboratory bilirubinometer SBR was 0.87 (0.77-0.93). The accuracy decreased with increasing haematocrit and at higher humidity ( $\geq 75\%$ ). Direct

## Open Peer Review

Reviewer Status 

	Invited Reviewers		
	1	2	3
<b>version 2</b> (revision) 23 Nov 2018			 report
<b>version 1</b> 04 Sep 2018	 report	 report	 report

1. **Rebecca R. Richards-Kortum**, Rice

University, Houston, USA

**Meaghan Bond**, Rice University, Houston, USA

2. **Robert D Christensen**, University of Utah, Salt Lake City, USA

observation of the operators using the device and analysis of the focus group discussions and questionnaires indicated that the BS was considered easy to use and required limited training.

**Conclusions:** This evaluation showed that the BS, in its current formulation, does not provide reliable results for measuring SBR in a tropical, low-resource setting but has acceptable usability features.

### Keywords

Neonatal hyperbilirubinemia, point-of-care, bilirubin, low resource setting

3. **Bolajoko O Olusanya** , Centre for Healthy Start Initiative (HSI-Centre), Lagos, Nigeria

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the [Mahidol Oxford](#)

[Tropical Medicine Research Unit \(MORU\)](#)

gateway.

**Corresponding author:** Laurence Thielemans ([thielemans.laurence@gmail.com](mailto:thielemans.laurence@gmail.com))

**Author roles:** **Thielemans L:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Hashmi A:** Formal Analysis, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Priscilla DD:** Investigation, Project Administration; **Kho Paw M:** Investigation, Project Administration; **Pimolsorntong T:** Investigation; **Ngerseng T:** Data Curation; **Van Overmeire B:** Supervision, Writing – Review & Editing; **Proux S:** Conceptualization, Funding Acquisition, Resources, Validation; **Nosten F:** Conceptualization, Funding Acquisition, Methodology, Supervision, Validation, Writing – Review & Editing; **McGready R:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Carrara VI:** Conceptualization, Methodology, Supervision, Writing – Review & Editing; **Bancone G:** Conceptualization, Funding Acquisition, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Wellcome Trust [106698], Major Overseas Programme–Thailand Unit, which supports the Shoklo Malaria Research Unit, part of the Mahidol Oxford University Research Unit; TL was supported by a PhD grant from 'The Belgian Kids' Fund for Pediatric Research'.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2018 Thielemans L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** 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 [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2018, 3:110 <https://doi.org/10.12688/wellcomeopenres.14767.1>

**First published:** 04 Sep 2018, 3:110 <https://doi.org/10.12688/wellcomeopenres.14767.1>

## Introduction

Neonatal hyperbilirubinemia (NH), a common disorder worldwide, has a benign course if promptly managed. However, in settings with limited access to diagnosis and care, there is a higher morbidity and mortality risk<sup>1</sup>. Accurate testing for elevated serum bilirubin (SBR) in the first days of life, followed by appropriate treatment with phototherapy, is crucial to prevent brain damage in infants<sup>2,3</sup>. Serum bilirubin concentration is the gold standard reference for determining NH management<sup>2-5</sup>, and it is done by laboratory-based colorimetric assay, which requires large blood volumes and a fully-equipped laboratory—often beyond the reach of clinics in low-resource settings. While visual assessment by Kramer zone scores<sup>6</sup> poorly correlates with bilirubin concentration<sup>7-9</sup>, transcutaneous bilirubinometers (TcB) have proven to be an alternative to invasive blood sampling<sup>10-12</sup>. However, TcB may underestimate high SBR and, as they assess extravascular bilirubin, are unreliable if the neonate has received phototherapy treatment<sup>13-15</sup>. The need for postnatal age- and ethnicity-specific nomograms further complicate TcB use<sup>16,17</sup>.

Frontline health workers in low-resource settings need reliable and cost-effective point-of-care (POC) tests for early screening of neonates at risk, timely treatment or referral when phototherapy is not locally available and monitoring of neonates requiring treatment. A promising, low-cost device called BiliSpec has been field tested in Malawi, showing good agreement with the standard reference (Pearson's correlation,  $r=0.97$ ) and a mean difference of  $5.1 \mu\text{mol/l}$  with 95% confidence interval between  $-29.1$  and  $37.6 \mu\text{mol/l}$ <sup>18</sup>. An alternative POC device is the Bilistick (BS) system, which features a hand-held, battery-operated machine requiring a smaller blood volume and generating a faster result than the standard laboratory test. When compared with laboratory assays on duplicate plasma samples of 118 neonates in Italy and Egypt under ideal conditions, BS showed an acceptable mean underestimation of SBR levels of  $10.3 \mu\text{mol/l}$  and a 95% confidence interval between  $-38.0$  and  $58.7 \mu\text{mol/l}$ <sup>19</sup>. In 2017, in a tertiary care referral centre in Cairo, samples from neonates with SBR less than  $340 \mu\text{mol/l}$  were used to investigate the accuracy of the BS on venous blood<sup>20</sup>. The study concluded that BS had comparable accuracy to the TcB JM-103. A recent multi-centre study reported the BS to slightly underestimate SBR value with maximum limits of agreements of  $-128.3$  to  $102.6 \mu\text{mol/l}$ <sup>21</sup>.

These findings suggest that new POC assays to diagnose NH have promise for use in low-resource settings, but there is insufficient evidence to make strong recommendations. The current study used a mixed-methods approach to evaluate two important features of the BS: (i) the technical accuracy of the BS in tropical field conditions, and (ii) usability of the device among locally trained staff in a low resource setting.

## Methods

### Outline and location

This prospective, mixed-methods study was conducted by the Shoklo Malaria Research Unit (SMRU) at the Wang Pha clinic, located along the Thailand-Myanmar border in the north-western province of Tak, Thailand. The study evaluated

BS during two distinct periods: from the end of January until the end of April 2017, and from the beginning of June until the end of December 2017.

The field clinic and on-site laboratories have stable electricity, basic equipment and refrigerators. The staff consists of locally trained medics, midwives, nurses, health workers and laboratory technicians. Prior to starting the study, all midwives ( $n=13$ ) and 3 laboratory technicians were asked and agreed to participate to the study. They were then trained on use of the BS system. The training module was created and conducted by the first author (L.T.) based on protocols developed by the manufacturer; user instructions were simplified and translated for the local staff. The training consisted of two sessions: 1 hour of theory with a step-by-step description of the procedure supported by pictures and videos, followed by 1 hour of practical, hands-on training.

The study included neonates born at  $\geq 35$  weeks of gestational age, in stable clinical condition, and requiring bilirubin measurement for either (i) a grade 3 or higher jaundice observed visually as per the Kramer scale<sup>6,8</sup>, or (ii) a previous, borderline SBR measurement (i.e.  $\leq 50 \mu\text{mol/l}$  below the treatment threshold of the British National Institute for Health and Care Excellence (NICE) guidelines<sup>3</sup>). All parents of neonates who met the inclusion criteria were proposed to participate in the study. The study was explained in the preferred local language and the parents who agreed in participating signed an informed consent form.

Since the BS required a haematocrit (HCT)  $\leq 65\%$  to provide a reading, neonates with a prior HCT  $>65\%$  or no prior HCT measured were not eligible. Each infant enrolled contributed to one measurement only.

Gestational age was defined by ultrasound<sup>22,23</sup>. Neonatal age in hours and phototherapy treatment at time of SBR measurement were reported systematically.

### Technical validation

Simultaneously, two blood samples were collected by heel prick: one ( $50 \mu\text{l}$ ) in a heparinized capillary tube for laboratory testing and one ( $25 \mu\text{l}$ ) in a non-heparinized transfer pipette for BS.

The first  $50\text{-}\mu\text{l}$  blood sample was used for the routine laboratory bilirubinometer SBR testing and transferred to the on-site clinic laboratory within 10 minutes of collection. The laboratory staff centrifuged the capillary tube (3 min at 10,000 rotations per minute) to separate red blood cells from plasma and estimate HCT using a Hawksley micro-haematocrit reader. To ensure quality, the accuracy of capillary HCT reading was assessed independently from this study. A total of 90 consecutive HCTs were independently read by two laboratory staff. The maximum observed difference was 11 HCT points; in 54% (48/89) of the samples, the two readings were the same, in 37% (33/89) the difference was 1 HCT point and in 9% (8/89) a difference of  $\geq 2$  HCT points was observed. There were no statistical differences between paired HCT (Wilcoxon matched pairs signed rank test,  $p=0.089$ ).

In the study, after HCT reading, the same tube was used to measure SBR using the dual-wavelength BR-501 bilirubinometer (Apel Co., Ltd, Japan) according to manufacturer's instructions. The BR-501 bilirubinometer has  $\pm 5\%$  accuracy within the measurement range (0–30 mg/dl), as reported by the manufacturer.

The second blood sample was used for the study device. The BS system (Bilimetrix srl, Italy) consists of a hand-held reflectance reader that uses test strips composed of a filter coupled with a nitrocellulose membrane ([Supplementary File 1](#) and <http://www.bilimetrix.net/>). The device is powered by rechargeable batteries. The BS calibration was checked weekly according to the manufacturer's instructions and kept under ambient conditions for the entire duration of the study. The test strips were kept in their humidity-proof packaging and were exposed to ambient conditions for a few seconds prior to performing the test.

The test was performed according to the manufacturer's instructions after computerized installation and registration. Although the reader was designed for a single operator, the midwives decided to perform the test in pairs. One midwife inserted a test strip in the BS system for calibration (38 sec to determine the reflectance of the dry strips) while the second midwife collected the two blood samples via heel prick. The blood was loaded on the strip within 2 minutes of instrument calibration. Measurement was displayed within 100 seconds ([Supplementary File 1](#)). If the blood was not loaded within 2 minutes of the calibration, the strip had to be removed and the reader required 2 additional minutes to reset prior to accepting a new strip. The new strip would require a new calibration and the total time would potentially cause clotting of the blood. After each test, a visual evaluation of the strip was performed by the operator to confirm saturation of the membrane. If the reader was unable to indicate an accurate result for the bilirubin level, it generated an error message displayed on the BS screen. As per approved protocol, neonates were not pricked a second time and the BS test was not repeated when clotting occurred or error outputs messages were displayed.

Temperature and room humidity were recorded with a temperature hygrometer at the time of the test. Humidity level was classified as either normal or high humidity ( $< 75\%$  or  $\geq 75\%$ , respectively).

### Usability assessment

The mixed-methods approach employed both exploratory and explanatory design and was used to evaluate the usability of the BS system. This included observation by laboratory technicians of technical procedures performed by the midwives and midwives' assessment of the device's ease of use through focus group discussions (FGD) and a self-administered questionnaire. The questionnaire and FGD took place after 6 months of using the BS. This mixed-methods approach also allowed for triangulation through independent qualitative analysis of FGD data by two investigators (L.T., A.H.), and mutual corroboration between observed technical errors and midwives' evaluation of the device's ease of use via questionnaire.

A total of three laboratory technicians evaluated each step of the process using an internal assessment checklist ([Supplementary File 2](#)) and reported technical issues (e.g., blood clotting).

After providing written informed consent, midwives filled in a 5-point Likert scale, self-administered questionnaire that was developed based on the operating procedures of the user manual ([Supplementary File 3](#)). The questions were translated into the local language (Sgaw Karen), back-translated, and finalized prior to conducting this component of the study. The 5-point Likert scale ranged from "Strongly disagree" (score 1) through "Neutral" (score 3) to "Strongly agree" (score 5).

After completing the questionnaire, midwives participated in FGD to explore their experiences working with the BS device ([Supplementary File 4](#)). FGD guides and questions were developed based on investigator knowledge and information shared by the manufacturer. FGD guides were translated into Sgaw Karen, back-translated and finalized prior to conducting FGD. To promote an environment in which participants felt comfortable talking, participants were purposively selected to form groups similar in clinical experience. There were two focus groups conducted, one with six junior and senior midwives and one with seven assistant midwives. The FGD facilitator (M.K.P.) was fluent in Sgaw Karen, the preferred language of the midwives, and assisted by L.T. FGD were audio-recorded and transcribed directly to English by a translator unfamiliar and uninvolved in the study.

### Data analysis

Sample size calculations for technical validity of BS assumed a power of 80%, an effect size of 0.25 and  $\alpha = 0.0125$ , yielding a sample size of at least 180 samples needed to detect a statistically significant difference in measurement compared to routine laboratory bilirubinometer via two-tailed paired t-test. Proportions were compared using chi-square test and Mann-Whitney U-test was used for comparison of medians. A Bland-Altman plot was used to graphically inspect differences between the two tests. NICE guidelines<sup>3</sup> were used for clinical interpretation of routine laboratory bilirubinometer SBR to diagnose NH and direct treatment as needed. Data were analysed using STATA/IC 14.0 (Stata Corp 2015, Version 14.1. College Station, Texas, StataCorp LP).

The results of FGD were analysed by two researchers (A.H., L.T.). L.T. reviewed the transcripts and clarified any outstanding questions with M.K.P. Initially, open coding was employed upon preliminary reading of the transcripts (L.T. and A.H.). Then, line-by-line inductive coding was performed by L.T. and A.H., independently. Mind maps were constructed to help reach consensus on codes and coding hierarchy. Each researcher independently performed thematic analysis based on initial coding to identify emergent themes. Discrepant coding between the two investigators was discussed until consensus reached and coding adjusted accordingly. Queries were then performed on the textual data to identify similarities and variations in the two FGD. Analysis was facilitated by NVivo for Mac v11.4.0 (QSR International, Melbourne, Australia).

In order to create an easy-to-use scale to analyse, two questions of the self-administered questionnaire (nr1 and nr9) had to be reversed. Scores were then described using median (IQR) and their distribution demonstrated by a histogram. Reliability of the questionnaire was assessed by the Cronbach alpha coefficient<sup>24,25</sup>.

## Ethical clearance

The study was approved by the Ethic Committees of the University of Oxford, UK (OXTREC 5115-16) and the Faculty of Tropical Medicine, Mahidol University, Thailand (TMEC16-073). The Tak Community Advisory Board, consisting of members of the local community, also revised and approved the study (TCAB-05/Rev/2016). All 13 midwives trained to perform the BS consented to the self-administered questionnaire and participate in FGD. Written informed consent was obtained from the parents or guardian of the neonates.

## Results

### Overview of validation data

Consent was obtained from parents of 180 eligible neonates. In total, three tests were excluded from analysis because the laboratory technicians were unavailable to supervise the procedure; four additional tests were excluded from further evaluation: three had insufficient amounts of blood to complete the BS test, and the reader stopped functioning during one test, yielding

an incomplete test result. Overall, BS was performed on 173 blood samples (Figure 1).

Most neonates enrolled in the study were born at term (162/173, 93.6%). The test was done at a median age of 4 days; the median (IQR) value of the last HCT prior to recruitment was 57 (54–61), ranging from 31% to 65% and the median (IQR) HCT value at recruitment was 55% (52–60). Eleven samples had a HCT value above 65%, the cut-off HCT to allow correct saturation of the BS strip membrane. Median room temperature at time of sampling was 28.9°C (range: 23–36°C) and humidity 67% (range: 39–89%). A total of eight neonates were under phototherapy treatment at the time of sampling (Table 1).

### Technical validation

There were two readers used over the course of this study (Supplementary File 5). An error message was generated by the reader for all the 11 samples above the HCT threshold of readability (>65%) and no SBR readout was available (Table 2).

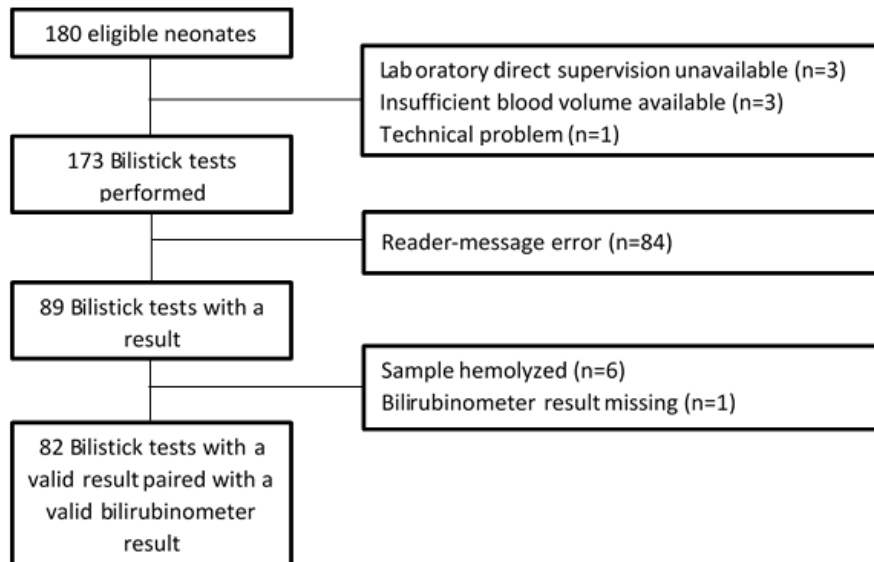


Figure 1. Study flowchart.

Table 1. Baseline newborn and environment characteristics at time of bilirubin measurement.

Characteristics*	All (n=173)	BS result (n=89)	BS errors (n=84)	p-value
<b>Newborn</b>				
Gestational age, weeks	39 [38-40]	39 [38-40]	39 [38-40]	0.948
Age, hours	99 [63-161]	128 [70-170]	83 [56-146]	0.008
Haematocrit, %	55 [52-60]	53 [49-55]	59 [56-63]	<0.001
Under phototherapy, n (%)	8 (4.6)	3 (3.4)	5 (6.0)	0.429
<b>Environmental</b>				
Room temperature, °C	29 [28-30]	29 [28-30]	29 [56-63]	0.245
Room humidity, %	67 [58-77]	69 [58-78]	67 [57-76]	0.489

\*Median [Interquartile range] unless otherwise stated.



**Table 2. Proportion of error messages by neonatal haematocrit (HCT) categories.**

HCT* range, %	All, n	Total errors, n (%)	Reader #1 (n=90)	Reader#2 (n=82)
31 ≤ 40	2	0 (0)	0/2 (0)	-
41 ≤ 55	87	19 (21.8)	9/42 (21.4)	10/45 (22.2)
56 ≤ 65	72	54 (75.0)	33/37 (89.2)	21/35 (60.0)
66 ≤ 75	11	11 (100.0)	9/9 (100.0)	2/2 (100.0)

\*1 missing HCT. Proportion of error messages increased significantly with higher HCT (Chi square for trend  $p < 0.001$  overall,  $p < 0.001$  with Reader #1 and  $p = 0.005$  with Reader #2).

Furthermore there was no SBR readout in 75% (54/72) samples with a HCT range of 56–65% and in 22% (19/87) samples with a HCT range of 41–55% (Table 2).

There were three types of reader-generated error messages encountered (Supplementary File 6); EC:T06 (Incomplete reading procedure within the established time), EC:B03 (Error identified during the initial phase of the bilirubin measurement) and EC:B04 (Error identified during the bilirubin measurement). The majority (64/84, 76.2%) were due to a failure of the reading procedure (described by the manufacturer as potentially linked to the HCT level). They were all found in equal proportion at different humidity levels (<75%, ≥75%,  $p = 0.242$ ), but the proportion of error “EC:B04” (i.e., insufficient test strip saturation) was significantly lower ( $p = 0.004$ ) with Reader #2 (Supplementary File 6). The proportion of reader-generated error messages did not differ significantly among different midwives ( $p = 0.306$ ) or over time (chi-square for trend per month,  $p = 0.392$ ) and was not associated with gestational age, room temperature or humidity (Table 1). Supplementary File 5 reports the history of reader-generated error messages and the technical support needed from manufacturers during the course of the study.

Of the 89 samples with an available SBR result, 7 were excluded from further analysis: the reader reported haemolysis for 6 tests and one had no laboratory bilirubin measured (Figure 1). The remaining 82 BS measurements could be compared with routine laboratory bilirubinometer SBR results: 36 from Reader #1 and 46 from Reader #2.

Reader #2 was mostly used during a period of high humidity (58.7% of the tests) and on samples with higher HCT (Table 3). The correlation between BS and routine laboratory bilirubinometer SBR was higher for Reader #1; the Pearson’s correlation coefficient ( $r$ ) was 0.97 (95% CI: 0.93–0.98) compared to 0.71 (95% CI: 0.53–0.83) for Reader #2 (Figure 2A, B). Taking the two readers together, this correlation varied depending on the ambient humidity; there was a high correlation at humidity levels < 75% with a Pearson’s correlation coefficient  $r = 0.96$  (95% CI: 0.94 to 0.98), but it decreased to 0.62 (95% CI: 0.34 to 0.80) when measurements were done at high levels of humidity (≥ 75%). The HCT also influenced the BS measurement accuracy with a correlation coefficient of 0.95 (95% CI: 0.91 to 0.97) for the HCT range of 31–55% and 0.48 (95% CI: –0.04 to 0.80) for the HCT of 56–65%.

When the humidity and HCT conditions were combined, the highest correlation was found for the HCT range 30–56% at humidity levels of < 75% ( $r = 0.97$  (95% CI: 0.88–1.00)) and the lowest correlation was at a HCT range of 56–65% at levels of high humidity ( $r = 0.29$  (95% CI: 0.03–0.65)) (Table 4).

The Bland–Altman plot of the BS measurements ( $\mu\text{mol/l}$ ) against routine laboratory bilirubinometer SBR performed at an ambient humidity of <75% ( $n = 52$ ) showed an acceptable mean difference of  $-20 \mu\text{mol/l}$  (limits of agreement  $-59$  to  $18$ ). The maximum observed difference was  $-74 \mu\text{mol/l}$  (Figure 3A). At ambient humidity levels ≥75% ( $n = 30$ ) there was a mean difference of  $-21 \mu\text{mol/l}$  with wide limit of agreement of  $-117$  to  $75$  and a high maximum difference of  $-211 \mu\text{mol/l}$  (Figure 3B).

#### Usability assessment

**Direct observation of 173 sample measurements.** Midwives were consistently able to insert the strip into the reader, draw blood with the pipette, load blood onto the strips and read the results. The laboratory technicians reported a total of 9 major (5.2%) and 3 minor (1.7%) inaccuracies while observing midwives performing the procedure (Supplementary File 2). The most frequent error was the formation of blood bubbles while loading the strip ( $n = 9$ ). All 12 observed errors happened within the first 3 months of the study (first month  $n = 5$ , second month  $n = 2$ , third month  $n = 2$ ); 10 of those errors led to a reader-generated error message (Supplementary File 2), while the remaining two gave valid results (no haemolysis was reported by the reader).

Blood clotting was reported in two instances when the blood was not loaded within 2 minutes of the strip calibration, leading to the need of performing a new calibration and thus a delay in reading. Both tests led to a reading error message (Supplementary File 6).

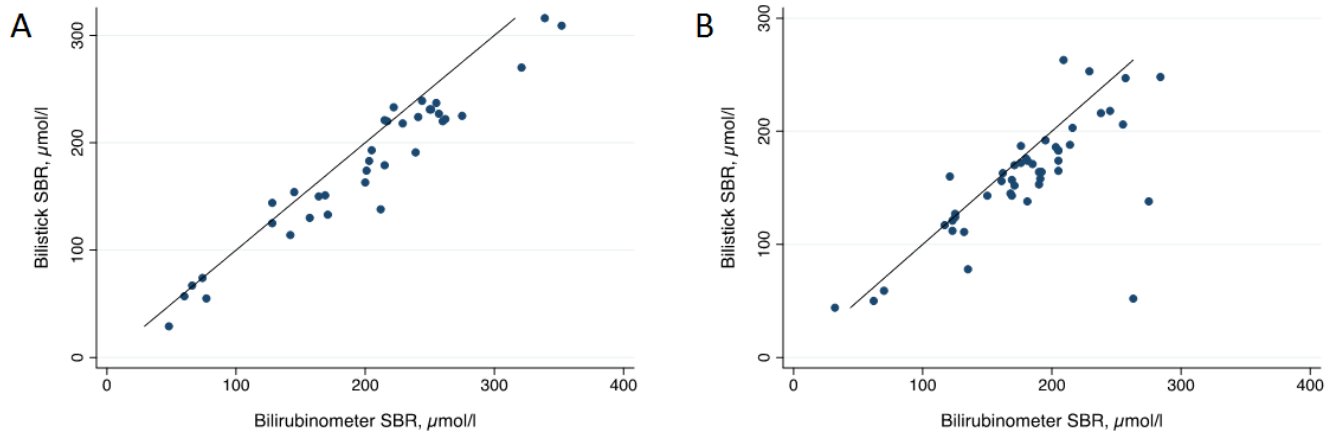
**Focus group discussion.** Emergent themes included: knowledge acquisition, midwife experience around training and use of BS; the clinical potential and considerations in implementing the BS at a larger scale; and midwife report of parents’ experiences.

Midwives knew the importance of screening for and diagnosing NH. The overall impression of midwives was that the training was simple, easy to understand, and greatly enhanced by the use of pictures. In fact, midwives thought that easy mastery of the instrument indicated that BS was appropriate to use in more

**Table 3.** Comparison of Bilistick System (BS) Reader#1 and #2 from available results (n=82).

Variables*	All (n=82)	Reader #1 (n=36)	Reader #2 (n=46)	p-value
<b>Newborn characteristics</b>				
Gestational age, weeks	39 [38-40]	39 [38-40]	39 [38-40]	0.359
Age, hours	127 [70-169]	127 [82-221]	125 [68-168]	0.824
Haematocrit, %	53 [49-55]	50 [49-54]	54 [52-56]	0.046
<b>Environment characteristics</b>				
Room temperature, °C	29 [28-30]	28 [27-31]	29 [29-30]	0.782
Room humidity, %	69 [58-78]	58 [52-65]	76 [69-81]	<0.001
<b>Humidity level, n (%)</b>				
Normal	52 (63.4)	33 (91.7)	19 (41.3)	
High	30 (36.6)	3 (8.3)	27 (58.7)	
<b>Bilirubin measurement characteristics</b>				
BS bilirubin, $\mu\text{mol/l}$	168 [138-218]	187 [136-226]	164 [138-186]	0.266
Bilirubinometer bilirubin, $\mu\text{mol/l}$	192 [150-229]	214 [151-251]	181 [150-205]	0.045
PCC of bilirubin values between BS and routine laboratory bilirubinometer, r (95% CI)	0.87 (0.77-0.93)	0.97 (0.93 - 0.98)	0.71 (0.53 - 0.83)	

\*Data presented as median [Interquartile range] unless otherwise stated. PCC, Perason's correlation coefficient.



**Figure 2.** Correlation of bilirubin measurement between BS and routine laboratory bilirubinometer. Solid line is the  $y=x$  reference line. (A) Reader #1; (B) Reader #2. SBR, serum bilirubin.

**Table 4.** Pearson correlation coefficient under different humidity and haematocrit levels.

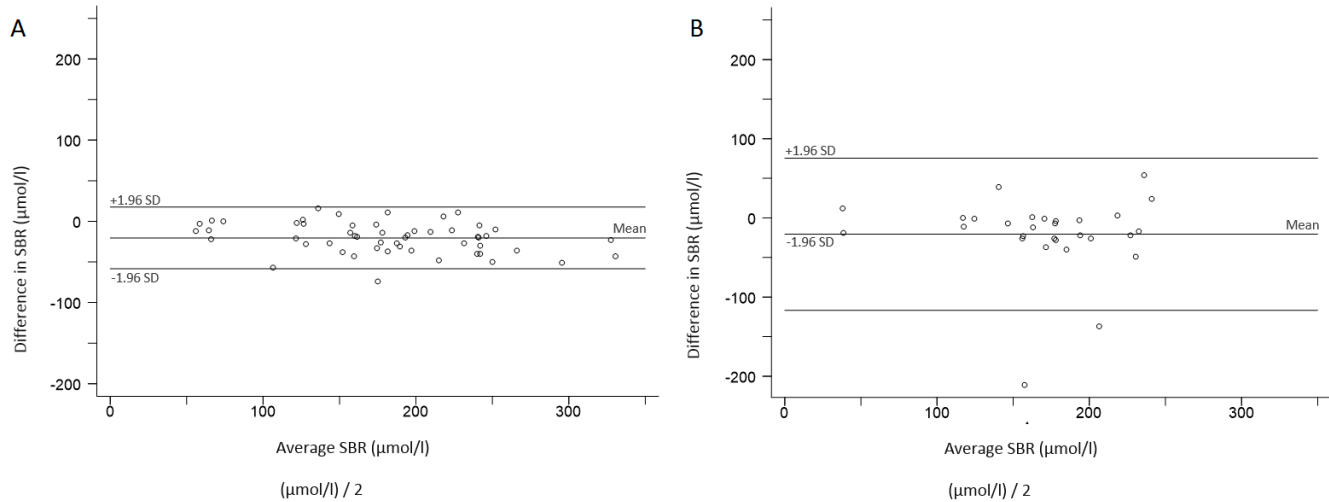
Haematocrit level	Humidity level, n; r (95%CI)	
	<75%	≥75%
30-55%	45; 0.97 [0.95-0.98]	22; 0.86 [0.69-0.94]
56-65%	7; 0.90 [0.48-0.99]	8; 0.29 [0.03-0.65]

rural and remote settings provided health workers received appropriate training.

*'If they [village health workers] are familiar and can use it well, they can use it. You can also train the health worker to use the device because it is easy to use.'*—Assistant midwife

However, the midwives expressed confusion by the way results were displayed (in two units:  $\mu\text{mol/l}$  and  $\text{mg/dl}$ ) as they were only familiar with the  $\mu\text{mol/l}$  results.

*'If the result comes up, it is not the same as the lab result... We know how to look at the lab result, but with the BS, how can we look at? ... How can we look for the result in using BS because there are many numbers?'*—Senior midwife



**Figure 3. Bland-Altman plot for Bilistick System BS results as compared to routine laboratory bilirubinometer results. (A)** Humidity <75% (n=52); **(B)** humidity  $\geq$ 75% (n=30). Difference in SBR = (Bilistick SBR – Bilirubinometer SBR) ( $\mu\text{mol/l}$ ), Average SBR = (Bilistick SBR + Bilirubinometer SBR) ( $\mu\text{mol/l}$ ) / 2. SBR, serum bilirubin.

Furthermore, midwives frequently commented on the device’s easiness of use. Another positive finding was the immediacy with which results could be obtained, often cited as a benefit compared to long wait times associated with the routine laboratory bilirubinometer testing.

*In my point of view, the BS is working faster than the machine used in the lab. When we send to the lab, we have to wait a little bit. And if there are a lot of patients and we are busy, we have to wait even longer.*—Junior midwife

However, the bulk of the discussion around general use of the device centred on the frequency of errors. Indeed, this reflected a widely-held frustration of midwives using this device.

*‘We do as she [L.T.] teaches us but the result doesn’t appear on the device. Sometimes the result appears on the device. So I have no idea with this—is it because of what we did or because of the device? I don’t understand.’*—Assistant midwife

In the context of result reliability, settings appropriate for BS use were frequently discussed and often reflected training considerations. Midwives asserted that frequent errors limited the use of the device to facilities with adequate laboratory support. However, if results could be ensured, the speed with which the BS device could produce an accurate SBR measurement made the device favourable for use in low resource settings.

*‘We can use it without needing a lab if error does not appear. We just worry: if an error appears, we’ll have to start [the procedure] again. Without a lab, how do we do? So we need a lab.’*—Assistant midwife

*‘If we use the lab, we have to look for the result. But the device has automatic answers appearing in numbers so that we don’t need to look for the result.’*—Senior midwife

Finally, midwives asserted that the speed of diagnosis translated into prompt treatment of NH, with clear benefit for patients in low resource settings.

*‘[The device is] good for the children. If the children need phototherapy, we can give them in time before it gets worse. We can prevent severity.’*—Senior midwife

Parental concern arose as an important theme in discussions with midwives. Midwives reported that parents expressed concern and fear of heel pricks that may cause the infant pain. Thus, midwives would worry about having to repeat the tests due to BS error and often had to spend time counselling parents on the importance of heel pricks or repeat heel pricks for accurate diagnosis of NH. One assistant midwife mentioned that the parents of a baby, having witnessed multiple heel pricks, became increasingly wary of causing the infant undue pain leading them to avoid even routine immunizations.

*‘[The parents] have compassion on the children when they see that the prick hurt their children. But parents don’t accept if [midwives] repeat the prick too many times. There’s a time that they tell us...It happens that they [parents] rejected vaccination because their children had been pricked so many times’*—Assistant midwife

**Usability questionnaire.** All thirteen midwives completed the questionnaire. [Supplementary File 7](#) shows the distribution of participant’s answers on a 5-point Likert scale. The midwives agreed on the ease of use, with a median (IQR) score of 4 (4-4). The two questions that rated low were the only two negative statements; one about the training, “I did not need to learn a lot of things before I could get going with the Bilistick system” and one on the ability of the midwife to perform the test independently, “I don’t need the support of a technical person (lab technician) to be able to use properly the Bilistick system”. The results on these two statements were not in line with the midwives’ experiences assessed by FGD



and it was suspected that the construction of the question as a negative was misunderstood. Cronbach alpha coefficient for Likert scale questions inter-reliability was low (0.5) with all statements included, but increased to acceptable internal reliability levels of 0.8 after removing the two negative statements. The median score remained the same after the two negative statements were removed.

## Discussion

This study evaluated the hand-held bilirubin reader BS system developed to measure and display SBR concentration, for technical accuracy and usability in a low-resource setting<sup>20</sup>.

The studied population was healthy near-term and term neonates in their first days of life (median age 4 days) with a high median capillary HCT of 55%. The young postnatal age and the use of capillary blood in the current study may explain the high HCT levels observed<sup>26,27</sup>. Moreover, delayed cord clamping is regularly practiced at the SMRU clinic according to WHO recommendations<sup>28</sup> resulting in newborns having a median HCT of 59% (IQR, 54–64) at 24 hours of life<sup>29</sup>, a value similar to other settings worldwide<sup>26,27,30</sup>.

This population contrasts with that of the three previous studies that tested the BS under ideal conditions: in the first study, neonates had a lower mean HCT value of 41.5%<sup>19</sup>, those enrolled in the second study were of an older median age of 6 days with venous blood<sup>20</sup> and in the most recent multi-centre study, the median HCT was 42.7% IQR (38.0–48.0) at a median age of 4 days<sup>21</sup>.

The current study detected a clear limitation of the readers in analysing SBR levels starting from HCT range 41–55% and major limitation in the HCT range 55–65%. Greco *et al.* documented a proportion of 6.8% (11/161) of technical problems using venous blood in older neonates<sup>20</sup>, while the current study experienced 48.6% of “reader-generated error messages”. Overall, the device in its current formulation would not be suitable for screening in the early hours of life when mean HCT often ranges from 55 to 60%<sup>26,27,30</sup> and in settings where capillary sampling is used.

For the tests that yielded a result, the accuracy of both BS readers with humidity <75% were similar to those already published, showing comparable measures of the BS system by non-invasive rapid transcutaneous bilirubinometers<sup>20</sup>. However, the BS accuracy was reduced with increasing HCT value (>55%) and higher humidity (≥75%). Had the BS test been used for NH diagnosis and management, only one neonate would have met the criteria for phototherapy according to NICE guidelines, compared to the five neonates diagnosed by routine laboratory bilirubinometer SBR. Given its lack of utility in younger neonates and its inaccuracy at high humidity and HCT levels, this study concludes that BS in its current configuration is not reliable for early screening of NH in this tropical setting.

This technical evaluation had some limitations: the BS was not compared to the gold standard, which would have required

a larger volume of blood with venous sampling and analysis at the central laboratory in Mae Sot (30 km away). Moreover, due to the limited number of strips available, the high rates of errors, and the use of two different BS readers, the number of samples analysed in similar environmental conditions was small. At the time of manuscript submission, the BS manufacturer has already changed some features of the test strips and reports proper performances with blood sample with HTC up to 70%; further investigation will be required to assess the performances of the new device.

Assessing usability through both exploratory (observation) and explanatory (FGD, questionnaire) methods proved a strong approach. It confirmed that not only was the absence of SBR readout frequent, but that it gave the user considerable anxiety. Importantly, direct observation by laboratory technicians highlighted the small number of errors performed by the midwives when performing the BS test. Therefore, the assertion by the midwives that the device was easy to use, required limited training, and could be mastered in more remote settings is valid. However, the data from the usability assessment support the technical validation: the system proved to be poorly suited for low resource settings<sup>31</sup>. More specific caveats in the use of this system require particular attention. The device required a level of technological capacity often beyond the means of low-resource settings<sup>31</sup>. For example, the device requires computerized installation, registration prior to use, and—if technical assistance is required—remote internet access with compatible software. In addition to these technological constraints, one must consider the local capacity in information technology support to troubleshoot as problems arise. Secondly, the test itself required metered blood collection with a specific pipette, prompt loading of the strip as the pipettes are non-heparinized, and recalibration between each sample. Very strict time constraints are impractical in settings where the operator needs to sample blood from a newborn in a busy clinic. Although midwives reported the test to be easy to use, they spontaneously started performing the test in pairs in spite of its single-operator design. In addition to these clinical concerns, perceived infant discomfort of heel pricks led some parents to be sceptical of SBR testing; repeated sampling in a short time, because of close monitoring for borderline SBR values, in apparently healthy babies can be perceived as very distressing to the parents. This indicates the need to find a less invasive bilirubin screening tool for this setting.

Improving the performance under high humidity and with higher HCT levels, possibly with an improved strip matrix, could potentially make the BS system suitable for tropical settings. Heparinized pipettes, longer times for loading the sample on the strip or alternatively, collecting blood from a heel prick directly onto the strip, may improve the usability of the system. Minor changes such as results being displayed for a longer period of time and options to select preferred measurement units would benefit the system.

In conclusion, despite its innovative approach and easiness to use, the BS tested at SMRU is not suitable for SBR measurement in tropical, low-resource settings.

### Data availability

Due to ethical and security considerations, the data that supports the findings in this study can be accessed only through the Data Access Committee at Mahidol Oxford Tropical Medicine Research Unit (MORU). The data sharing policy can be found here: <http://www.tropmedres.ac/data-sharing>. The application form for datasets under the custodianship of MORU Tropical Network can be found in [Supplementary File 8](#).

---

### Grant information

This work was supported by the Wellcome Trust [106698], Major Overseas Programme–Thailand Unit, which supports

the Shoklo Malaria Research Unit, part of the Mahidol Oxford University Research Unit; TL was supported by a PhD grant from ‘The Belgian Kids’ Fund for Pediatric Research’.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

### Acknowledgements

The authors would like to thank the mothers and babies participating in the cohort and all the staff of SMRU clinics and laboratories for their hard work and dedication. They also thank the Bilimetrix team for lending the Bilistick device.

### Supplementary material

Supplementary File 1. Schematic representation of the (A1) Bilistick System and (A2) Bilistick display. (*Figures have been reproduced with permission from the Bilistick user manual*).

[Click here to access the data.](#)

Supplementary File 2. Procedural errors made by midwives under observation by the supervising laboratory technician. An internal Bilistick assessment checklist was used.

[Click here to access the data.](#)

Supplementary File 3. The 5-point Likert scale self-administered questionnaire for midwives to assess the easiness of use of the Bilistick System.

[Click here to access the data.](#)

Supplementary File 4. Focus group guide used with midwives to assess the usability of the Bilistick.

[Click here to access the data.](#)

Supplementary File 5. History of reader generated error messages and technical support.

[Click here to access the data.](#)

Supplementary File 6. Numbers, types and descriptions of BS generated errors accrued.

[Click here to access the data.](#)

Supplementary File 7. Distribution of 13 participants’ answers on the easiness of use of the Bilistick System.

[Click here to access the data.](#)

Supplementary File 8. Application form required to access raw data from this study.

[Click here to access the data.](#)

---

### 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.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maisels MJ, Bhutani VK, Bogen D, *et al.*: **Hyperbilirubinemia in the newborn infant > or =35 weeks' gestation: an update with clarifications.** *Pediatrics.* 2009; 124(4): 1193–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- National Institute for Health and Care Excellence (NICE): **Neonatal Jaundice, clinical guidelines.** <http://www.nice.org.uk>. 2010.  
[Reference Source](#)
- WHO: **Pocket book of Hospital care for children: Guidelines for the Management of Common Childhood Illnesses.** 2nd ed. Geneva; 2013.  
[Reference Source](#)
- Raimondi F, Ferrara T, Borrelli AC, *et al.*: **Neonatal hyperbilirubinemia: a critical appraisal of current guidelines and evidence.** *J Pediatr Neonatal Individ Med.*

- 2012; **1**(1): 25–32.  
[Publisher Full Text](#)
6. Kramer L: **Advancement of dermal icterus in the jaundiced newborn.** *Am J Dis Child.* 1969; **118**(3): 454–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  7. Moyer VA, Ahn C, Sneed S: **Accuracy of clinical judgment in neonatal jaundice.** *Arch Pediatr Adolesc Med.* American Medical Association. 2000; **154**(4): 391–4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  8. Szabo P, Wolf M, Bucher HU, *et al.*: **Detection of hyperbilirubinaemia in jaundiced full-term neonates by eye or by bilirubinometer?** *Eur J Pediatr.* 2004; **163**(12): 722–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  9. Keren R, Tremont K, Luan X, *et al.*: **Visual assessment of jaundice in term and late preterm infants.** *Arch Dis Child Fetal Neonatal Ed.* 2009; **94**(5): F317–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  10. Yu ZB, Han SP, Chen C: **Bilirubin nomograms for identification of neonatal hyperbilirubinemia in healthy term and late-preterm infants: a systematic review and meta-analysis.** *World J Pediatr.* Childrens Hospital, Zhejiang University School of Medicine. 2014; **10**(3): 211–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  11. Nagar G, Vandermeer B, Campbell S, *et al.*: **Reliability of transcutaneous bilirubin devices in preterm infants: a systematic review.** *Pediatrics.* 2013; **132**(5): 871–81.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  12. Bhutani VK, Gourley GR, Adler S, *et al.*: **Noninvasive measurement of total serum bilirubin in a multiracial predischARGE newborn population to assess the risk of severe hyperbilirubinemia.** *Pediatrics.* American Academy of Pediatrics. 2000; **106**(2): E17.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  13. Taylor JA, Burgos AE, Flaherman V, *et al.*: **Discrepancies between transcutaneous and serum bilirubin measurements.** *Pediatrics.* American Academy of Pediatrics. 2015; **135**(2): 224–31.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  14. Bosschaart N, Kok JH, Newsum AM, *et al.*: **Limitations and opportunities of transcutaneous bilirubin measurements.** *Pediatrics.* 2012; **129**(4): 689–94.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  15. Nagar G, Vandermeer B, Campbell S, *et al.*: **Effect of Phototherapy on the Reliability of Transcutaneous Bilirubin Devices in Term and Near-Term Infants: A Systematic Review and Meta-Analysis.** *Neonatology.* Karger Publishers. 2016; **109**(3): 203–12.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  16. Wainer S, Rabi Y, Parmar SM, *et al.*: **Impact of skin tone on the performance of a transcutaneous jaundice meter.** *Acta Paediatr.* 2009; **98**(12): 1909–15.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  17. Samiee-Zafarghandy S, Feberova J, Williams K, *et al.*: **Influence of skin colour on diagnostic accuracy of the jaundice meter JM 103 in newborns.** *Arch Dis Child Fetal Neonatal Ed.* 2014; **99**(6): F480–4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  18. 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.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  19. Coda Zabetta CD, Iskander IF, Greco C, *et al.*: **Billistick: a low-cost point-of-care system to measure total plasma bilirubin.** *Neonatology.* 2013; **103**(3): 177–81.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  20. 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.* Nature Publishing Group. 2017; **37**(9): 1028–1031.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  21. Greco C, Iskander IF, El Houchi SZ, *et al.*: **Diagnostic Performance Analysis of the Point-of-Care Bilistick System in Identifying Severe Neonatal Hyperbilirubinemia by a Multi-Country Approach.** *EClinicalMedicine.* The Author(s). 2018; **1**: 14–20.  
[Publisher Full Text](#)
  22. Papageorgiou AT, Kennedy SH, Salomon LJ, *et al.*: **International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester of pregnancy.** *Ultrasound Obstet Gynecol.* 2014; **44**(6): 641–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  23. Rijken MJ, Lee SJ, Boel ME, *et al.*: **Obstetric ultrasound scanning by local health workers in a refugee camp on the Thai-Burmese border.** *Ultrasound Obstet Gynecol.* 2009; **34**(4): 395–403.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  24. Rattray J, Jones MC: **Essential elements of questionnaire design and development.** *J Clin Nurs.* 2007; **16**(2): 234–43.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  25. Sullivan GM, Artino AR Jr: **Analyzing and interpreting data from likert-type scales.** *J Grad Med Educ.* 2013; **5**(4): 541–2.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  26. Jopling J, Henry E, Wiedmeier SE, *et al.*: **Reference ranges for hematocrit and blood hemoglobin concentration during the neonatal period: data from a multihospital health care system.** *Pediatrics.* 2009; **123**(2): e333.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  27. Gatti RA: **Hematocrit values of capillary blood in the newborn infant.** *J Pediatr.* Elsevier. 1967; **70**(1): 117–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  28. WHO: **Guideline: Delayed umbilical cord clamping for improved maternal and infant health and nutrition outcomes.**  
[Reference Source](#)
  29. Thielemans L, Trip-Hoving M, Bancone G, *et al.*: **Neonatal Hyperbilirubinemia in a Marginalized Population on the Thai-Myanmar Border: A study protocol.** *BMC Pediatr.* 2017; **17**(1): 32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  30. Kalaivani K: **Prevalence & consequences of anaemia in pregnancy.** *Indian J Med Res.* 2009; **130**(5): 627–33.  
[PubMed Abstract](#)
  31. Engel N, Ganesh G, Patil M, *et al.*: **Point-of-care testing in India: missed opportunities to realize the true potential of point-of-care testing programs.** *BMC Health Serv Res.* BioMed Central. 2015; **15**: 550.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:   

---

## Version 1

Reviewer Report 21 September 2018

<https://doi.org/10.21956/wellcomeopenres.16093.r33822>

© 2018 Olusanya B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Bolajoko O Olusanya** 

Centre for Healthy Start Initiative (HSI-Centre), Lagos, Nigeria

### General

The authors of this paper rightly recognised the need for reliable, cost-effective and easy to use point of care (POC) bilirubin measuring devices in developing countries. They set out to determine the reliability and usability of a newly developed POC device, Bilistick System (BS) at a field clinic along the Thailand-Myanmar border region using a mixed-methods approach from January to December 2017. The main findings were that the BS did not produce usable results in almost half of the tests but was found to be easy-to-use by the midwives. The authors thus concluded that BS was easy-to-use but not reliable for measuring serum bilirubin in a low-resource setting.

### Specific issues to address

The introduction to the study and the references cited throughout the paper clearly reflect the authors' lack of familiarity with the relevant studies from low and middle-income countries (LMICs). For example, the authors only recognised that TCB underestimates serum bilirubin whereas there is preponderance of evidence that TCB more commonly overestimates serum bilirubin in special populations in LMICs. Studies that seek to address needs in LMICs must acknowledge and build on the body of evidence from the region to be considered as a credible and valuable addition to knowledge.

Generally, the methodology is satisfactory. However, the authors need to clarify if the study participants were all in-born as the article seems to suggest. How were the out-born neonates managed? How was gestational age determined in these babies? How much did it cost to test a neonate using BS?

The conclusions are valid. BS in its current form does not appear to provide any benefit over the non-invasive TCB devices as its reliability diminishes with increasing haematocrit level above 55%. Perhaps, more crucially, the study has highlighted a major drawback and context for evaluating a recent multi-centre study<sup>1</sup> that seems to have exaggerated the utility of BS in developing countries. Whereas Greco et al reported technical problems in only 6.8% of the tests, the current study showed a significantly higher percentage (48.6%) of error messages. Evidently,

improvements in the BS are warranted.

The study reported that the 82 paired results from BS and bilirubinometer were analysed. The authors however, reported that a limitation of the study was that BS was not compared to the gold standard. The authors need to clarify which gold standard is required other than serum bilirubin.

Abbreviations used should be summarised at the end of the manuscript.

## References

1. Greco C, Iskander I, El Houchi S, Rohsiswatmo R, et al.: 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 [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Partly

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 07 Nov 2018

**Laurence Thielemans**, Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand

Please find below our responses to the comments of the reviewer.

1. The introduction has been modified as suggested, including additional references to TCB performances in LMICs.



2.1. Being born outside SMRU clinics was not an exclusion criterion. Mothers who delivered outside (i.e. at home) were advised to visit SMRU clinics as soon as possible after the delivery and all neonates were managed in the same way at the clinics.

2.2. Gestational age was routinely estimated either by ultrasound at the first antenatal consultation (ANC) or by Dubowitz gestational assessment at birth. The information was added in the methods and results.

2.3 We did not perform a formal cost analysis for testing BS in our setting. For the study, the Bilistick system was lent by Bilimetrix together with 200 strips and pipettes. Information was added in the methods.

3. For assessment of serum bilirubin, the laboratory gold standard is analysis by high performance liquid chromatography (HPLC, Kazmierczak et al., 2002) but spectrophotometric assays are more commonly used and considered reliable as reference tests. For this study the serum bilirubin was measured with the dual-wavelength BR-501 bilirubinometer (Apel Co., Ltd, Japan). We have clarified this point in the discussion.

4. A summary of the used abbreviations has been added at the end of the manuscript as suggested

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 12 September 2018

<https://doi.org/10.21956/wellcomeopenres.16093.r33830>

© 2018 Christensen R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Robert D Christensen**

Division of Neonatology, University of Utah, Salt Lake City, UT, USA

Thielemans *et al.* report a prospective investigation of the feasibility and diagnostic accuracy of a potential method for quantifying serum bilirubin in a tropical, low-resource, setting. The device is called the Bilistick System, and involves a small quantity of blood, typically drawn by heel-stick, and run as in a battery-operated, point of care system. These features would render this method far more practical and economical than standard laboratory-based serum bilirubin testing.

Severe hyperbilirubinemia that goes undetected and untreated can lead to acute bilirubin encephalopathy, when bilirubin crosses from blood to brain and damages, sometimes permanently, the central nervous system. The consequent death or life-long handicapping conditions, often consisting of deafness and athetoid cerebral palsy, place a large burden on families and societies. This tragic condition is one of the most heart-wrenching medical safety concerns in global pediatrics, because the damage is, in theory, completely preventable with early

recognition and treatment of neonatal hyperbilirubinemia.

The device did not test sufficiently well as to recommend its use in present configuration. However, the study appears to have been well conceived and executed and may give others the focus needed to improve the performance of this system, particularly in conditions with hematocrits above 55% and humidity condition above about 75%.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Neonatal Hematology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 11 September 2018

<https://doi.org/10.21956/wellcomeopenres.16093.r33824>

© 2018 Richards-Kortum R et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Rebecca R. Richards-Kortum**

Department of Bioengineering, Rice University, Houston, TX, USA

**Meaghan Bond**

Rice University, Houston, TX, USA

This interesting paper describes a well-designed study to evaluate the accuracy and usability of a

new point-of-care test to measure total serum bilirubin in low-resource settings. The results indicate an unacceptably high number of errors, despite observation of target users that indicated the test was being performed properly. The results will be of interest to researchers in the global health and the diagnostics communities. In particular, the paper highlights previously unreported challenges associated with achieving accurate results in high humidity and for neonates with high hematocrit levels.

The authors are encouraged to consider a number of minor questions and comments that would be of interest or helpful to readers:

1. Please include the author-developed training materials as Supplemental Information. What are the primary differences between the manufacturer and the author-developed training materials? Could this have affected results?
2. The authors state that "Since the BS required a haematocrit (HCT)  $\leq 65\%$  to provide a reading, neonates with a prior HCT  $>65\%$  or no prior HCT measured were not eligible."
  1. How many babies does this exclude?
  2. Why were babies with no prior HCT measured not eligible, since HCT was measured for the 50 uL gold standard blood sample?
3. It would be helpful to clarify the calibration process for the device. The methods state that the calibration was "checked weekly," but also that the user must insert a strip a maximum of 2 minutes before each sample for calibration.
4. To facilitate comparison of data from the two readers, it would be more helpful to put data from both readers on same graph with different symbols used for each.
5. It would be helpful to clarify earlier in paper that clotting due to delays occurs in the non-heparinized blood collection pipette, not on the BS strip.
6. The data in supplemental 5 (details about malfunctions and the manufacturer's attempts to fix them) seem important for real-world use and for understanding the high percentage of error messages. Could the authors move this paragraph to the main body of the paper.
7. The observations of how the test was used in practice are interesting. Could the authors briefly describe how the 2-minute time limit between strip calibration and sample measurement monitored. Is this actively measured by the BiliStick device? If not, how can the user tell that 2 minutes has passed and the sample is no longer accurate?
8. The data in Figure 2 show differences in the performance of the two readers; based on data in Tables 3 and 4, these differences are attributed to differences in the humidity at the time of use and the fraction of samples with high HCT measured using reader 2. Are there sufficient data to allow the authors to compare results for Readers 1 and 2 under the ideal conditions of humidity  $<75\%$  and HCT between 30-45% and assess whether there are any remaining reader-to-reader variations?

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** We have developed a similar point-of-care test to measure total serum bilirubin (reference 18 in the paper). We do not have a financial interest in this test.

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 07 Nov 2018

**Laurence Thielemans**, Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand

We appreciate the comments of the reviewer Rebecca R. Richards-Kortum. Please find below our Point-by-point answers.

1. The training material has been added as a supplementary file as suggested. The manufacturer provided a user manual, a user instruction poster and a video in English. However they were too difficult (and the video also too rapid) to understand for non-native English speaking staff and the instruction poster and user manual contained an error (acknowledged by the manufacturer). A simplified power point presentation was used instead, explaining the procedure step-by-step, highlighting the importance of timing at each step (turning the device on in time, waiting for the reader to display an instruction, inserting the strip in time etc.). Additional topics were addressed such as how to load the pipette correctly, how to deal with time constraints or a slow blood flow, and how to check the BS strip after the test. The required weekly calibration procedure was also explained. The video was replaced by a step-by-step demonstration on-site.

As we did not perform a training using the original material it is not possible to quantify or assess the impact but it is unlikely that adapting the training material affected the results negatively; indeed simplified explanations and live demonstration are usually expected to improve user understanding of procedures. The content of the training was the same as the instruction manual and poster, only its form was adapted. In essence, this was a process of tailoring instructions for appropriate BS use to the local context.

2.1. We did not record the number of non-eligible neonates in this study. We can estimate

that around 20-25% of children seen within 24h of life might have been excluded in this setting, based on results from a recently completed analysis birth cohort (ref 29 in manuscript doi: 10.1186/s12887-017-0798-8), where capillary Hct was measured routinely between 12-30h of life; the proportion of neonates with capillary Hct >65% was 17.4% (193/1111) and with capillary Hct  $\geq$ 70% 8.4% (93/1111) (unpublished data). We have added a sentence in the discussion in line with this (see paragraph 4, last sentence).

2.2. The Hct reading from the standard blood sample was not readily available (due to a 30 minute delay) and we did not want to heel prick the neonates twice. This, combined with a high proportion of neonates with Hct >65% and a small number of available strips (only 200 for a sample size of 180), led to the decision to maximize the number of valid readings and thus exclude babies without prior known Hct values.

3. We have now clarified the complete calibration process of the BS device in the methods.

4. Figure 2 was modified as recommended.

5. We have now clarified this in the methods.

6. The paragraph has been added to the main manuscript (in the result: technical support) as suggested.

7. The 2-minute time limit was calculated by the Bilistick device itself and a warning message was displayed once the time limit was reached so that a new strip had to be inserted. We clarified the process in the methods.

8. The analysable data performed under the ideal conditions of humidity <75% and HCT 30-55% are limited to 30 and 15 tests with Reader #1 and Reader #2, respectively. Pearson correlation coefficient was similar in both groups ( $r=0.97$  (95% CI: 0.93–0.98) and  $r=0.98$  (95% CI: 0.94–0.99)). The information has been added in the results.

**Competing Interests:** No competing interests were disclosed.