A Novel Icterometer for Hyperbilirubinemia Screening in Low-Resource Settings

Anne CC Lee, MD, MPH,[a] Lian V. Folger, BA,[a] Mahmoodur Rahman, MBBS,[b,c] Salahuddin Ahmed, MBBS,[c] Nazmun Nahar Bably, MPH, MBBS,[c,d] Lauren Schaeffer, MSc,[a] Rachel Whelan, MSc,[a] Pratik Panchal, MD, MPH,[a] Sayedur Rahman, MBBS,[c] Arun Dutta Roy, MBBS,[c] Abdullah H. Baqui, MBBS, MPH, DrPH[c]

abstract

BACKGROUND: Severe neonatal hyperbilirubinemia (>20 mg/dL) affects ~1 million infants annually. Improved jaundice screening in low-income countries is needed to prevent bilirubin encephalopathy and mortality.

METHODS: The Bili-ruler is an icterometer for the assessment of neonatal jaundice that was designed by using advanced digital color processing. A total of 790 newborns were enrolled in a validation study at Brigham and Women's Hospital (Boston) and Sylhet Osmani Medical College Hospital (Sylhet, Bangladesh). Independent Bili-ruler measurements were made and compared with reference standard transcutaneous bilirubin (TcB) and total serum bilirubin (TSB) concentrations.

RESULTS: Bili-ruler scores on the nose were correlated with TcB and TSB levels (r = 0.76 and 0.78, respectively). The Bili-ruler distinguished different clinical thresholds of hyperbilirubinemia, defined by TcB, with high sensitivity and specificity (score ≥3.5: 90.1% [95% confidence interval (CI): 84.8%–95.4%] and 85.9% [95% CI: 83.2%–88.6%], respectively, for TcB ≥13 mg/dL). The Bili-ruler also performed reasonably well compared to TSB (score ≥3.5: sensitivity 84.5% [95% CI: 79.1%–90.3%] and specificity 83.2% [95% CI: 76.1%–90.3%] for TSB ≥11 mg/dL). Areas under the receiver operating characteristic curve for identifying TcB ≥11, ≥13, and ≥15 were 0.92, 0.93, and 0.94, respectively, and 0.90, 0.87, and 0.86 for identifying TSB ≥11, ≥13, and ≥15. Interrater reliability was high; 97% of scores by independent readers fell within 1 score of one another (N = 88).

CONCLUSIONS: The Bili-ruler is a low-cost, noninvasive tool with high diagnostic accuracy for neonatal jaundice screening. This device may be used to improve referrals from community or peripheral health centers to higher-level facilities with capacity for bilirubin testing and/or phototherapy.

[a]*Global Newborn Health Laboratory, Department of Pediatric Newborn Medicine, Brigham and Women's Hospital and Harvard Medical School, Harvard University, Boston, Massachusetts;* [b]*Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh;* [c]*Department of International Health, International Center for Maternal and Newborn Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland; and* [d]*Institute of Epidemiology, Disease Control, and Research, Dhaka, Bangladesh*

Dr Lee conceptualized and designed the Bili-ruler and validation study, coordinated and supervised data collection and analysis, and drafted, reviewed, revised, and finalized the manuscript; Ms Folger implemented the study, collected data at Brigham and Women's Hospital, conducted icterometer assessments, coordinated the study, conducted data analysis for both sites, and reviewed and revised the manuscript; Dr M. Rahman helped design and implement the study in Sylhet, Bangladesh, conducted icterometer assessments, entered data, conducted data analysis, and reviewed the manuscript; Drs Ahmed, Bably, S. Rahman, and Roy helped implement the study in Sylhet, Bangladesh, and reviewed the manuscript; Ms Schaeffer and Ms Whelan (Continued)

WHAT'S KNOWN ON THIS SUBJECT: Severe neonatal hyperbilirubinemia affects an estimated 1 million infants annually. The majority of bilirubin-related encephalopathy and mortality is concentrated in low-income countries, where accurate and low-cost methods for jaundice screening are urgently needed.

WHAT THIS STUDY ADDS: The Bili-ruler, a novel, low-cost icterometer, can be used to classify different clinical thresholds of hyperbilirubinemia with high diagnostic accuracy. This low-cost tool may facilitate more accurate and timely neonatal jaundice management by frontline health workers in low-income settings.

To cite: Lee AC, Folger LV, Rahman M, et al. A Novel Icterometer for Hyperbilirubinemia Screening in Low-Resource Settings. *Pediatrics.* 2019;143(5):e20182039

Clinical jaundice is highly prevalent in the newborn period, affecting up to 84% of healthy newborns ≥35 weeks' gestation in the first week of life.[1,2] Annually, severe hyperbilirubinemia, defined as total serum bilirubin (TSB) >20 mg/dL, affects 1.1 million infants, and extreme hyperbilirubinemia (TSB >25 mg/dL) affects 481 000 infants.[3] In high-income countries, kernicterus has virtually been eliminated with increased access to early screening and treatment with high-intensity phototherapy.[4–7] However, in low- and middle-income countries (LMICs), access to screening, monitoring, and treatment of hyperbilirubinemia is limited. In these settings, an estimated 6 million infants who need phototherapy do not receive it[8]; approximately one-third of infants with extreme hyperbilirubinemia die, and 44% develop severe encephalopathy.[3]

The first barrier in the management of neonatal hyperbilirubinemia is the delay in identifying the problem and seeking care. Lack of access to the measurement of bilirubin concentration remains an obstacle in most LMICs. Serum bilirubin laboratory testing and transcutaneous bilirubin (TcB) devices are frequently unavailable or too costly. In the current World Health Organization Integrated Management of Childhood Illness guidelines, frontline health workers rely on visual inspection to refer newborns whenever jaundice is identified at <24 hours of life or at any time if seen on the palms or soles of the feet.[9] However, by the time plantar jaundice is observed, bilirubin levels are substantially elevated (>20 mg/dL). Furthermore, clinical assessment is complicated by subjectivity and interobserver variability, and poorly correlates with serum bilirubin,[2] particularly in darkly pigmented infants.[6]

A simple and low-cost method to improve the objectivity of visual inspection for jaundice was originally developed by Thomas Gosset. The Gosset icterometer[10] (1954) was a handheld device with 5 shades of yellow color that was matched to the infant's blanched skin and used to approximate bilirubin levels. The device had good correlation with serum bilirubin concentrations and was highly sensitive (93%–100%) for detecting significant hyperbilirubinemia in studies of varying ethnicities, with lower specificity (56%–74%).[8–10] The icterometer was never widely adopted because of high cost, increasing availability of serum bilirubin testing, and limited testing in dark-skinned infants, and it is no longer manufactured or distributed. Another disadvantage of the original icterometer was the limited technology available at the time, because the original device was made from "tins of robbialac paint made up in 1954. . .; the colour shades of icterometers made from different batches of paint may not be identical."[10] To address some of these limitations, we designed a new icterometer using advanced digital color processing technology and an enhanced visual design to simplify color matching. The tool is aimed for use by frontline health workers in low-resource settings to screen for clinically significant hyperbilirubinemia requiring referral and treatment in a health facility.

In this study, we aimed to determine (1) the validity of the novel icterometer (Bili-ruler) to detect clinically significant thresholds of hyperbilirubinemia and (2) the interrater reliability of icterometer scoring.

METHODS

The Standards for Reporting Diagnostic Accuracy (STARD) 2015 checklist[11] is available in the Supplemental Information. The full protocol for this prospective study is available at request from the corresponding author.

Icterometer Design

We designed a new icterometer, the Bili-ruler, using advanced digital color processing techniques and visual, human-centered design to address limitations of the original Gosset icterometer. The Bili-ruler is constructed from digitally standardized and calibrated archival-quality paper color strips of increasing yellow hue, numbered 1 through 6 (Fig 1). The Bili-ruler prototypes used in this study were manufactured in house using a semiflexible acrylic base and adherent acrylic film to secure the color strips to the base. The Bili-ruler is multiuse and can be sanitized between patients.

To create the color strips, we obtained images of blanched skin of infants from Bangladesh with known serum bilirubin levels using an X-rite ColorChecker Passport (X-rite, MI) to (1) create a digital camera calibration profile and (2) standardize the color output for processing. Standardized color palettes, such as the ColorChecker Passport, are used in digital photography to neutralize the effect of the illumination on the output. We compiled a library of digital photographs of infants with varying levels of hyperbilirubinemia, ranging from none to severe, and also incorporated the highest score of the

FIGURE 1
Use of the Bili-ruler on the nose of an infant at BWH in Boston.

Gosset icterometer. A stepwise gradient was used to generate the color scale. The digital workflow and image processing were done in the sRGB 16-bit color space using Adobe Photoshop and Adobe Lightroom applications.

Another concern with the original icterometer was the challenge in decision-making with respect to color matching. To address this, instead of choosing between 2 linear color strips, we used a clear, circular window surrounded by a single uniform color swatch that requires the user to make a relatively simpler binary (yes or no) matching decision. A similar visual design has been used for easier color matching in anemia screening by Hemochek (Allied Health Sciences Pvt. Ltd, New Delhi, India), a tool based on the World Health Organization's Hemoglobin Color Scale.[12,13]

Study Participants

Newborns were recruited from Brigham and Women's Hospital (BWH), Boston, Massachusetts, and Sylhet Osmani Medical College Hospital (SOMCH), Sylhet, Bangladesh. Newborns were eligible for inclusion in the study if the infant was <28 days old. Infants were excluded if they had received phototherapy, exchange transfusion, or if the hospital physician deemed the infant too ill (<2000 g, very preterm, serious illness). At BWH, newborns were recruited from the well-newborn nursery. At SOMCH, newborns were recruited from the obstetrics department (labor and delivery ward) and the pediatric inpatient department. A study physician or research assistant assessed infants for eligibility and obtained signed consent from parents.

Bili-ruler Use

Newborns were assessed for jaundice using the Bili-ruler, without previous knowledge of transcutaneous or serum bilirubin levels. To use the Bili-ruler, starting at the first color strip, the user applies firm pressure to blanch the infant's skin and visualize the hue of the underlying subcutaneous tissue. This process is repeated for each color on the ruler (ranging 1–6), and the user chooses the Bili-ruler score that most closely matches the underlying skin color. Measurements were performed with natural light, typically near a window and without fluorescent lighting, as possible. Bili-ruler measurements were performed on the forehead, nose, abdomen, palms, and soles. Two independent Bili-ruler measurements were performed on each body part by a study physician or research assistant, without knowledge of reference bilirubin levels. The averages of the 2 scores were used in analysis.

For a random subset of infants enrolled in the study, 2 independent measurements on the nose by different examiners were performed to assess interrater agreement. Each examiner was blinded to the newborn's reference standard bilirubin levels and to the other examiner's Bili-ruler scores.

Reference Standard Bilirubin Measurements

A TcB level was obtained within 2 hours of Bili-ruler measurement by using a Drager JM-105 device on all infants enrolled. Serum bilirubin values were only included in the analysis if drawn within 2 hours of the Bili-ruler measurement. At BWH, all infants received routine TcB screening as part of clinical care, and serum bilirubin was only obtained per the hospital's clinical protocol (ie, if the TcB level was above a prespecified threshold dependent on the infant's postnatal age). At SOMCH, TcB screening was not performed as part of standard clinical care, and serum bilirubin was routinely used to guide clinical management. Blood was collected for TSB determination per hospital routine and before initiation of phototherapy or exchange transfusion. Serum bilirubin analysis was performed in Popular Laboratory in Sylhet, Bangladesh (Vitros 50600 machine), which performed daily calibration and quality control testing. SOMCH allowed the introduction of TcB measurement for all infants enrolled in the research study. The study physician shared TcB and TSB results with the SOMCH hospital physician caring for the infant.

Statistical Analysis

Stata 15.1 (Stata Corp, College Station, TX) was used for analyses. Simple descriptive statistics were used to calculate means, medians, and ranges of study population characteristics, as well as to determine mean and SD bilirubin values for each icterometer score. Participants missing either a Bili-ruler measurement or a reference standard bilirubin test were excluded from analysis. Spearman correlation coefficients were calculated to assess correlation of Bili-ruler scores and bilirubin values. We were interested in determining the diagnostic accuracy of the Bili-ruler to identify different clinical thresholds of hyperbilirubinemia (≥11, 13, 15, 17, 20 mg/dL). Sensitivity, specificity, positive predictive values (PPVs), negative predictive values (NPVs), and positive and negative likelihood ratios, with 95% confidence intervals (CIs), were calculated for different hyperbilirubinemia thresholds. Receiver operating characteristic (ROC) curves were generated, and areas under the ROC curve were calculated. Exploratory subgroup analyses by ethnicity were also conducted. Bland Altman plots were generated to compare agreement between TSB and TcB. The total sample size was calculated to estimate sensitivity and specificity with a precision of 6.5%, assuming an expected sensitivity and specificity of

85% and hyperbilirubinemia prevalence of 15%.

Ethics Statement

The study was approved by the Partners HealthCare Institutional Review Board (Boston, MA) and the Ethic Review Committee of the Bangladesh Medical Research Council (Dhaka, Bangladesh). The funding source was not involved in the design, conduct, analysis, interpretation, and writing up of the study, nor the decision to submit it for publication.

RESULTS

From March 2016 to June 2017, 790 neonates were enrolled at BWH (Boston) and SOMCH (Sylhet, Bangladesh). A diagram of participant enrollment is available in Supplemental Fig 4. Characteristics of the study population are shown in Table 1. There were differences between the populations of the 2 study sites because of the differences in study setting and recruitment. Mothers at BWH were older than those at SOMCH. Few infants at SOMCH had known birth weights or gestational ages, given low rates of antenatal care and facility delivery. Newborns recruited at BWH were generally well, recruited postpartum, and the proportion with elevated bilirubin levels was lower than at SOMCH, where more newborns were recruited at pediatric admission. No adverse events resulted from study interventions.

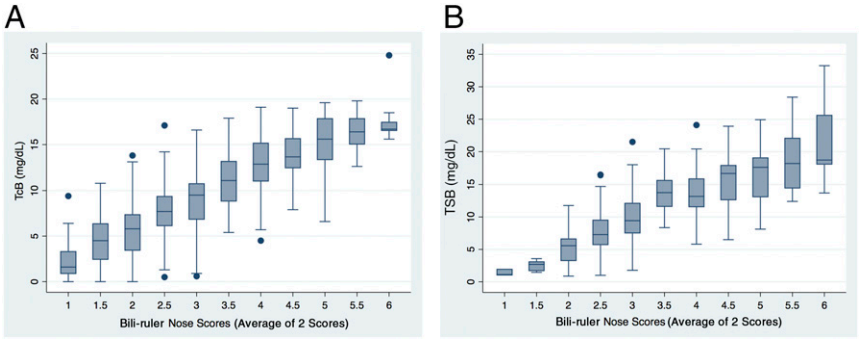

FIGURE 2
Boxplots of Bili-ruler scores on the nose compared with reference standard bilirubin levels. A, Bili-ruler scores versus TcB. B, Bili-ruler scores versus TSB.

Correlation of Bili-ruler Measurements and Bilirubin Concentrations

Overall, there was a strong positive correlation of icterometer measures on the nose with bilirubin concentrations (correlation coefficient: TcB, $r = 0.760$; TSB, $r = 0.780$; Fig 2 A and B). Measurements on the foot were also highly correlated with TSB levels ($r = 0.702$; Supplemental Fig 5B). Icterometer measures on the abdomen were only moderately correlated with TcB levels and TSB levels (TcB, $r = 0.532$; TSB, $r = 0.619$; Supplemental Fig 5 C and D). Skin on the abdomen was more difficult to blanch and visualize the underlying skin tone.

TcB and TSB were highly correlated ($r = 0.943$; $n = 235$) with one another, and the mean difference between

TABLE 1 Characteristics of Study Population

	BWH	SOMCH
Sample size	390	400
Ethnicity, No. (%)		
Asian	53 out of 390 (13.6)	400 out of 400 (100)
Black or African American	63 out of 390 (16.2)	0
Hispanic or Latino	37 out of 390 (9.5)	0
Non-Hispanic white	228 out of 390 (58.5)	0
Other or unknown	9 out of 390 (2.3)	0
Average maternal age (SD), y	33.1 (4.4)	23.5 (4.3)
Male, No. (%)	189 out of 390 (48.5)	263 out of 400 (65.8)
Preterm, <37 wk, No. (%)	19 out of 390 (4.9)	4 out of 5 (80)[a]
Low birth wt, No. (%)	6 out of 390 (1.5)	2 out of 14 (14.3)[a]
TcB available, No. (%)	390 out of 390 (100)	370 out of 400 (92.5)
Mean TcB level (SD)[b]	7.63 (3.4)	9.13 (5.3)
Median TcB level (range)[b]	7.7 (0–18.5)	9.1 (0–24.8)
TcB >15, No. (%)	8 out of 390 (2.1)	68 out of 370 (18.4)
TSB available, No. (%)	0	265 out of 400 (66.3)
Mean TSB (SD)[b]	—	12.2 (6.2)
Median TSB (range)[b]	—	12.4 (0.9–33.3)
TSB >15, No. (%)	—	83 out of 265 (31.3)

—, not applicable.
[a] A total of 395 SOMCH subjects were missing gestational age data; 386 SOMCH subjects were missing birth wt data.
[b] Unit for all bilirubin levels is mg/dL.

TcB-TSB levels was −0.31 mg/dL (SD: 1.8). A Bland Altman plot is shown in Supplemental Fig 6. There was no systematic bias, and the 95% limits of agreement were −3.79 to 3.16 mg/dL.

Mean Bilirubin Concentrations Associated With Bili-ruler Scores

In Table 2, we show the mean and SD TcB and TSB levels associated with Bili-ruler scores on the nose and foot. Given the cephalo-caudal progression of hyperbilirubinemia, we expected bilirubin levels to be more elevated when jaundice had reached the more caudal regions. This is consistent with our findings in Table 2, in which it is shown that a Bili-ruler score on the foot was associated with significantly higher bilirubin levels than those associated with the same Bili-ruler score on the nose.

Validity to Detect Different Hyperbilirubinemia Thresholds

In Table 3, we show the sensitivity, specificity, PPV, and NPV of the Bili-ruler to detect different thresholds of hyperbilirubinemia, with TcB and TSB references. The bilirubin thresholds were chosen on the basis of the treatment guidelines for low-income countries recommended by Bhutani et al[14] in 2011 and the American Academy of Pediatrics exchange transfusion thresholds.[15]

The Bili-ruler thresholds displayed in Table 3 were chosen to optimize both sensitivity and specificity. In Supplemental Table 6, we also show Bili-ruler thresholds at which 100% sensitivity is reached, to inform the clinician of the threshold at which a case of hyperbilirubinemia (for each cutoff) would not be missed (ie, 0% false-negative rate).

Compared with the TcB reference, Bili-ruler scores on the nose had high validity for identifying hyperbilirubinemia (Table 3, Supplemental Table 6). A score of 3 or higher had 91.3% (95% CI: 87.3%–95.2%) sensitivity and 74.5% (95% CI: 70.9%–78.1%) specificity for identifying a TcB ≥11 mg/dL. A score ≥3.5 had 90.1% (95% CI: 84.8%–95.4%) sensitivity and 85.9% (95% CI: 83.2%–88.6%) specificity for identifying a TcB ≥13 and 94.7% (95% CI: 89.6%–99.8%) sensitivity and 81.3% (95% CI: 78.4%–84.2%) specificity for identifying a TcB ≥15. A score of 4 or higher had 94.1% (95% CI: 86.2%–100%) sensitivity and 85.0% (95% CI: 82.2%–87.5%) specificity for identifying a TcB ≥17. Of note, the Drager device does not register a measurement at an equivalent serum bilirubin concentration of >20 mg/dL; the device displays an error signal and does not provide a numerical reading. Thus, we were unable to perform analysis for a TcB cutoff of ≥20 m/dL.

Results of the performance of the Bili-ruler compared to TSB are shown in Table 3 and Supplemental Table 7. A nose score ≥3.5 had 84.7% (95% CI: 79.1%–90.3%) sensitivity and 83.2% (95% CI: 76.1%–90.3%) specificity for identifying a TSB ≥11 mg/dL. A score ≥4 had 81.9% (95% CI: 75.1%–89%) sensitivity and 74.3% (95% CI: 67.3%–81.4%) specificity for identifying TSB ≥13 and 87.8% (95% CI: 80.9%–95.0%) sensitivity and 66.5% (95% CI: 59.6%–73.3%) specificity for identifying a TSB ≥15. Lastly, a score of ≥4.5 had 81.4% (95% CI: 71.4%–91.3%) sensitivity and 83.4% (95% CI: 78.3%–88.5%) specificity for identifying TSB ≥17 and 84.6% (95% CI: 70.7%–98.5%) sensitivity and 74.8% (95% CI: 69.3%–80.3%) specificity for identifying a TSB ≥20.

Bili-ruler scores on the foot of ≥2 had high sensitivity and lower specificity to identify a serum bilirubin ≥20 mg/dL (sensitivity 88.5% [95% CI: 76.2%–100%], specificity 77.0% [95% CI: 71.7%–82.3%]; Table 3).

ROC Curves

In Table 4, we show the area under the curve (AUC) for the ROC curves generated for Bili-ruler scoring on the

TABLE 2 Transcutaneous and Serum Bilirubin Levels Associated With Bili-Ruler Scores (Average of 2 Scores)

Bili-ruler Score[a]	Nose				Foot			
	TcB (*N* = 759)		TSB (*N* = 264)		TcB (*N* = 726)		TSB (*N* = 265)	
	n	mg/dL (SD)	*n*	mg/dL (SD)	*n*	mg/dL (SD)	*n*	mg/dL (SD)
1	16	2.4 (2.6)	3	1.4 (0.6)	509	7.2 (3.9)	160	9.0 (4.6)
1.5	51	4.4 (2.7)	5	2.5 (0.9)	97	9.5 (4.0)	27	14.5 (3.6)
2	227	5.5 (2.8)	40	5.4 (2.7)	109	12.4 (4.2)	59	16.7 (3.8)
2.5	143	7.6 (2.7)	21	8.0 (4.1)	3	8.6 (6.5)	3	18.0 (2.3)
3	123	8.8 (3.2)	44	10.0 (3.9)	6	15.0 (2.7)	12	20.9 (4.7)
3.5	58	11.3 (3.1)	18	13.8 (3.2)	0	—	1	33.3
4	66	13.0 (3.1)	51	13.7 (3.4)	1	18.5	1	23.0
4.5	26	13.8 (2.6)	23	15.7 (4.1)	0	—	0	—
5	31	15.3 (3.1)	32	16.6 (4.2)	0	—	1	30.0
5.5	9	16.3 (2.3)	12	18.8 (5.3)	1	24.8	1	32.3
6	9	17.7 (2.8)	15	21.6 (6.1)	0	—	0	—

—, not applicable.

[a] The Bili-ruler has 6 possible discrete scores, ranging from 1 to 6 (see Fig 1); the scores presented in this table are the average of 2 independent readings, thus allowing for half scores.

TABLE 3 Validity of Bili-Ruler Nose and Foot Measurements to Identify Clinically Relevant Thresholds of Hyperbilirubinemia Using Transcutaneous and TSB Reference Standards

Bilirubin Threshold, mg/dL	Bili-ruler Score Threshold[a] (Measurement Location)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)
TcB							
≥11	≥3 (nose)	91.3 (87.3–95.2)	74.5 (70.9–78.1)	55.3 (49.8–60.7)	96.1 (94.3–97.9)	3.58 (3.09–4.14)	0.117 (0.074–0.185)
≥13	≥3.5 (nose)	90.1 (84.8–95.4)	85.9 (83.2–88.6)	54.8 (47.9–61.7)	97.9 (96.7–99.1)	6.39 (5.23–7.80)	0.116 (0.067–0.198)
≥15	≥3.5 (nose)	94.7 (89.6–99.8)	81.3 (78.4–84.2)	35.7 (29.0–42.3)	99.3 (98.6–100)	5.06 (4.29–5.97)	0.066 (0.025–0.170)
≥17	≥4 (nose)	94.1 (86.2–100)	85.0 (82.2–87.5)	22.5 (15.7–29.4)	99.7 (99.2–100)	6.26 (5.13–7.52)	0.069 (0.018–0.266)
TSB							
≥11	≥3.5 (nose)	84.7 (79.1–90.3)	83.2 (76.1–90.3)	88.1 (82.9–93.2)	78.8 (71.2–86.3)	5.04 (3.29–7.71)	0.184 (0.126–0.268)
≥13	≥4 (nose)	81.9 (75.1–89)	74.3 (67.3–81.4)	71.6 (64.0–79.3)	84.0 (77.7–90.3)	3.19 (2.4–4.26)	0.244 (0.162–0.360)
≥15	≥4 (nose)	87.8 (80.9–95)	66.5 (59.6–73.3)	54.5 (46.0–62.9)	92.4 (87.8–96.9)	2.62 (2.11–3.27)	0.183 (0.100–0.327)
≥17	≥4.5 (nose)	81.4 (71.4–91.3)	83.4 (78.3–88.5)	58.5 (47.9–69.2)	94.0 (90.5–97.4)	4.91 (3.53–6.83)	0.224 (0.131–0.382)
≥20	≥4.5 (nose)	84.6 (70.7–98.5)	74.8 (69.3–80.3)	21.7 (17.2–36.4)	99.0 (95.7–99.9)	3.36 (2.55–4.41)	0.206 (0.083–0.508)
≥20	≥2 (foot)	88.5 (76.2–100)	77.0 (71.7–82.3)	29.5 (19.4–39.6)	98.4 (96.6–100)	3.84 (2.93–5.04)	0.150 (0.052–0.435)

[a] The Bili-ruler has 6 possible discrete scores, ranging from 1 to 6; the scores used in the analyses for this table were the average of 2 independent readings, thus allowing for half scores.

nose and foot for 5 bilirubin thresholds (≥11, 13, 15, 17, 20 mg/dL). In Fig 3, we show the ROC curve plot for Bili-ruler nose scores at different TcB thresholds and Bili-ruler foot scores with a TSB threshold of ≥20 mg/dL.

Bili-ruler scores on the nose had high AUCs to detect TcB ≥11, ≥13, ≥15, and ≥17 (all >0.90) (Table 4, Fig 3 A and B). AUCs to detect TSB ≥11, ≥13, ≥15, ≥17, and ≥20 were also high, ranging 0.85 to 0.90 (Table 4). The accuracy of foot scores compared with a serum bilirubin reference for the 5 bilirubin thresholds was moderate to high, with AUCs of 0.79 to 0.90 (Table 4, Fig 3C).

Stratified Analysis by Ethnicity

In Table 5, we show the mean TcB levels for each Bili-ruler score on the nose for the 4 major ethnic groups in our sample (black or African American, Hispanic-Latino, Asian, non-Hispanic white). Mean TcB values were within 2 mg/dL across ethnicities for all Bili-ruler scores with adequate sample size (>10 measurements). We did not have adequate TSB data to stratify analysis by ethnicity.

Interrater Reliability

Ninety-seven percent of the 88 Bili-ruler measurements by 2

TABLE 4 Areas Under the ROC Curve for Bili-Ruler Scoring on Different Body Sites

Bilirubin Cutoff (mg/dL)	Bilirubin Reference Standard	AUC for Nose Measurement	AUC for Foot Measurement
≥11	TcB	0.916	0.716
	TSB	0.896	0.786
≥13	TcB	0.930	0.753
	TSB	0.865	0.813
≥15	TcB	0.937	0.775
	TSB	0.859	0.833
≥17	TcB	0.939	0.802
	TSB	0.887	0.877
≥20	TSB[a]	0.848	0.900

[a] AUCs for bilirubin cutoff of 20 mg/dL are not shown for the TcB reference because the TcB machine would not register most bilirubin levels above 20 mg/dL (only 1 subject registered a TcB >20).

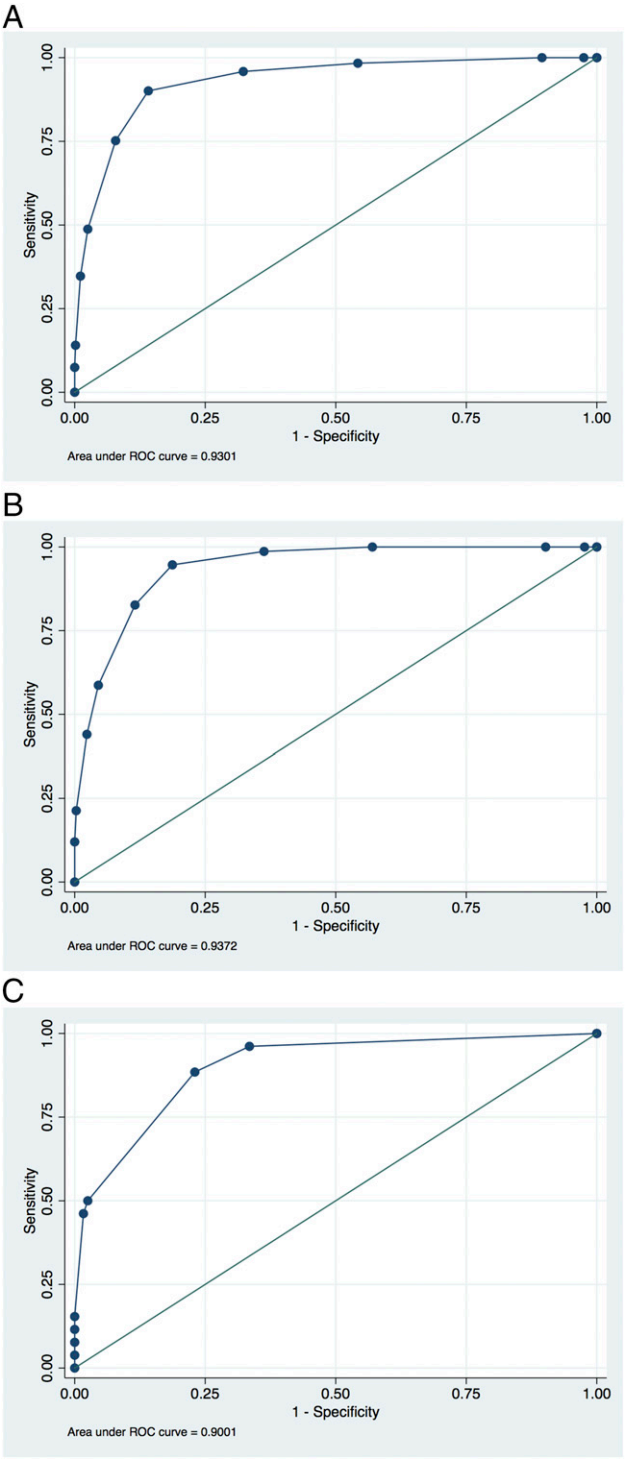

FIGURE 3

ROC curves for Bili-ruler scoring to identify clinically relevant bilirubin thresholds, defined by TcB and TSB. A, ROC curve for Bili-ruler scores on nose to identify TcB ≥13 mg/dL. B, ROC curve for Bili-ruler scores on nose to identify TcB ≥15 mg/dL. C, ROC curve for Bili-ruler scores on foot to identify TSB ≥20 mg/dL.

independent readers performed on the infants' noses were within 1 point (Supplemental Table 8). Agreement was within 0.5 points for 81% of interrater measurements on the nose.

DISCUSSION

We designed the Bili-ruler using advanced digital color processing and human-centered design to address limitations of previous devices. The Bili-ruler is low cost, portable, easy to color match, digitally standardized and calibrated, and accurate in identifying different clinically relevant thresholds of hyperbilirubinemia. We conducted a validation study including a multiethnic cohort of 790 newborns in Sylhet, Bangladesh, and Boston. Bili-ruler measures were highly correlated with TcB and TSB. The Bili-ruler distinguished clinically relevant thresholds of hyperbilirubinemia with high diagnostic accuracy. This low-cost, portable, and accurate tool may be used to improve the identification of hyperbilirubinemia in settings where laboratory testing and/or electronic transcutaneous devices are unavailable, and thus help to guide appropriate referral and case management.

Icterometry is an indirect visual measure of the transcutaneous degree of jaundice, indicating bilirubin concentrations in subcutaneous tissue and fat. The Bili-ruler performed better compared to a TcB reference versus a TSB reference. TcB and TSB are highly correlated with one another, although TcB levels may lag behind rapid rises in serum bilirubin that are seen in hemolytic diseases of the newborn.[16] The Better Outcomes through Research for Newborns (BORN) network collected data from 27 nursery sites in the United States and found a mean TcB-TSB difference of 0.84 ± 1.78 (SD) mg/dL,[17] and the investigators concluded that TcB served as a robust and reasonably accurate tool for TSB screening. In our study, the mean difference between TcB-TSB concentrations was 0.03 ± 1.95 (SD) mg/dL (n = 284). Given the noninvasive nature of TcB devices, they have replaced serum bilirubin for routine screening in

TABLE 5 Mean TcB Levels Associated With Bili-Ruler Scores on the Nose, by Ethnicity

Bili-Ruler Score (Nose)	Black or African American (*N* = 63)		Hispanic or Latino (*N* = 37)		Asian (*N* = 453)		Non-Hispanic White (*N* = 228)	
	n	Mean TcB in mg/dL (SD)	*n*	Mean TcB in mg/dL (SD)	*n*	Mean TcB in mg/dL (SD)	*n*	Mean TcB in mg/dL (SD)
1	2	1.0 (0.2)	0	—	3	1.1 (1.0)	11	3.1 (2.9)
1.5	6	6.1 (3.5)	4	3.5 (2.3)	9	4.5 (2.7)	31	4.2 (2.7)
2	14	5.9 (2.2)	19	6.2 (2.5)	111	4.8 (3.0)	80	6.3 (2.5)
2.5	18	9.0 (3.0)	10	7.5 (1.8)	63	6.9 (3.0)	51	8.1 (2.1)
3	14	9.9 (3.2)	2	8.9 (2.3)	74	8.2 (3.5)	29	9.7 (2.5)
3.5	7	11.3 (3.5)	1	13.0	32	11.9 (3.2)	18	10.1 (2.8)
4	0	—	1	11.2	61	13.1 (3.1)	4	12.1 (2.8)
4.5	0	—	0	—	22	13.9 (2.8)	4	13.2 (2.0)
5	2	17.1 (2.1)	0	—	29	15.2 (3.1)	0	—
5.5	0	—	0	—	9	16.3 (2.3)	0	—
6	0	—	0	—	9	17.7 (2.8)	0	—

—, not applicable.

many institutions in high-income settings where these devices are affordable. In LMICs, where the cost of transcutaneous devices is prohibitive, icterometry may thus serve as a useful tool for bilirubin screening.

In several previous studies, researchers have validated the original 5-color Gosset icterometer (also referred to as the Ingram icterometer) compared with TSB. In a primarily white population in Wisconsin, a score of ≥3 on the Ingram icterometer had sensitivity of 94% and specificity of 78% for identifying TSB >12.9 mg/dL.[18] In a Turkish study, an icterometer score >3 had high sensitivity (100%) yet lower specificity (48%) for identifying TSB >13.[19] A 2-color icterometer (Bilistrip) was developed in Nigeria for home use by mothers.[20] Whereas the performance for identification of TSB >12 mg/dL was fair (sensitivity 65%, specificity 90%), the predictive performance for a threshold of >15 mg/dL was good (sensitivity 85%, specificity 88%). Our new icterometer has 6 colors and is digitally standardized and designed for visual ease of color matching for frontline health workers in low-income settings. The performance of our icterometer to detect TSB was similar or slightly better than ranges reported in previous studies: A score of ≥3.5 on the nose had sensitivity of 84.7% (95% CI: 79.1%–90.3%) and specificity of 83.2% (95% CI: 76.1%–90.3%) to detect TSB ≥11 mg/dL (AUC: 0.90), and a score of ≥4 showed similar accuracy for detecting TSB of ≥13 and ≥15 mg/dL (AUCs: ≥13: 0.87; ≥15: 0.86). The diagnostic accuracy to detect TSB ≥20, an important threshold for American Academy of Pediatrics guidelines, was moderate (score of ≥2 on the foot: sensitivity 88.5%, specificity 77.0%; AUC: 0.90). Compared with a TcB reference, the diagnostic accuracy of the Bili-ruler was good to excellent, with sensitivity and specificity of 90.1% (95% CI: 84.8%–95.4%) and 85.9% (95% CI: 83.2%–88.6%), respectively, for a score of ≥3.5 to detect TcB ≥13 mg/dL and AUCs >0.90 for detecting all TcB thresholds tested.

Validity in varying ethnicities and skin tones is an important consideration for the use of icterometry in different populations. In our study, mean bilirubin levels for each icterometer score were relatively similar between various ethnic groups, with mean bilirubin within 2 mg/dL for Bili-ruler scores with adequate sample size. The largest populations validated in our study were the South Asian (*n* = 417) and white (*n* = 269) populations. The Ingram icterometer had previously been validated in different ethnicities, and, in a South Asian population, an icterometer score of 3 had 97% sensitivity and 71% specificity for identifying bilirubin >10 mg/dL in term infants.[21] In Tanzania, Hamel[22] found that icterometer readings on the gums were highly correlated with serum bilirubin (*R* = 0.91). Chaibva et al[23] validated the 1960 Gosset icterometer in a South African population and found that icterometer scores correlated highly with serum bilirubin measurements in infants with dark skin. However, serum bilirubin levels were slightly lower for each grade comparing dark- to light-skinned infants. Other studies have shown the TcB measures in African populations may have different validity than other ethnicities. In a study of neonates in Nigeria, TcB (Drager and Bilicheck) and TSB were highly correlated; however, TcB systematically overestimated TSB on average, by 2 to 3 mg/dL.[24] In the BORN study in the United States, the individual TcB-TSB differences were on average 0.67 mg/dL higher among African American neonates compared with neonates of other races.[17] In this study, we validated the icterometer in 63 African American infants. A larger validation study in African neonates is needed and currently being planned.

Given the low cost and ease of use, icterometry has potential for

community-based screening to identify at-risk infants in LMICs or even within ambulatory settings within high-income countries where transcutaneous devices are not available. An estimated one-half of births occurring in South Asia and Sub-Saharan Africa occur at home, and an estimated 27% of newborns in these regions have a first postnatal visit within 48 hours of life.[25] In these settings, community-based health workers or health extension workers may be the first line of contact. In current Integrated Management of Childhood Illness guidelines,[9] infants are referred for evaluation when they exhibit jaundice in the palms or soles, when it is often too late for rescue interventions. Thus, community-based screening and early detection of jaundice may lead to early evaluation and initiation of phototherapy treatment before levels reach exchange threshold. In several studies, researchers have tested the use of home-based icterometry by parents. In a pilot randomized controlled trial in Vietnam, home use of icterometers by mothers was feasible and helped increase detection of jaundice that was missed by visual inspection alone.[26] In Nigeria, a 2-colored icterometer (Bilistrip) was used by mothers for home-based monitoring.[20]

Understanding potential barriers to uptake and scale are important given that icterometers were previously available and considered valid. It is unclear whether the high cost or low availability of devices were reasons for low uptake in LMICs. Previous icterometers cost ~$15 per device in the 1950s, equating to ~$50 today with inflation. The Bili-ruler currently costs ~$5 per device; however, at scale, the ruler would cost <$1 per device to produce, which could improve potential for scalability. The current price point assumes materials cost at relatively low-volume production. We are exploring alternative manufacturing methods to support a more robust version that is suitable for large-scale production.

Our study had several limitations. In Bangladesh, we more heavily recruited from primarily hospital readmissions to the pediatric ward. Thus, this was a more selected population with higher rates of hyperbilirubinemia and may not reflect performance in the general population. We recruited a fair number of infants of dark skin tone in our Boston cohort (African American); however, the numbers were small (n = 63), and larger numbers are required for validation in different skin tones. We recruited primarily term infants and had a low number of preterm infants. In 1 study in India, the performance in preterm infants had somewhat lower sensitivity.[21] Given that the skin composition may differ in preterm infants, the icterometer should be validated in preterm populations. We also did not have sufficient data to validate the performance in older postnatal ages (ie, >7 days of life).

CONCLUSIONS

In low-resource areas, the availability and cost of serum bilirubin measurement are major obstacles to identifying newborns who require referral and/or phototherapy for the management of neonatal jaundice. The Bili-ruler addresses several of the limitations of the original Gosset icterometer and has excellent validity for identifying several different thresholds of hyperbilirubinemia. The Bili-ruler may enable health workers to more rapidly and accurately identify infants with hyperbilirubinemia at peripheral levels of the health care system or in communities, and provide them with early referral and/or timely treatment with phototherapy. This is the first step required to help reduce mortality and morbidity related to hyperbilirubinemia in the hardest-to-reach communities.

ACKNOWLEDGMENTS

We acknowledge Professor Manajiir and the staff at SOMCH for their support of the study. We also thank the staff and faculty at the BWH Department of Newborn Medicine for their support of this work, including Dr Terrie Inder, Dr Lise Johnson, Megan Dietrich, and the BWH Well-Newborn Nursery physicians and nurses. We thank Caroline Lau and Deborah Liu for their advice on the design, production, and distribution of the Bili-ruler, and Kristine Bunker and Sarah Folger for their input on materials and design. We acknowledge Dr Rich Fletcher for inputs on the visual design of the ruler and Dr Vinod Bhutani for early advice on the concept. We thank Shoshana Burgett and Arthur Schmehling of X-Rite Pantone for their input on design, manufacturing, and color measurement. We are grateful for the Saving Lives at Birth and Accelerator teams for their continued support of this work. Finally, this work would not be possible without the parents and infants who participated in this study.

ABBREVIATIONS

AUC: area under the curve
BWH: Brigham and Women's Hospital
CI: confidence interval
LMIC: low- and middle-income country
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
SOMCH: Sylhet Osmani Medical College Hospital
TcB: transcutaneous bilirubin
TSB: total serum bilirubin

helped implement the study, conduct icterometer assessments, enter and analyze data at Brigham and Women's Hospital, and reviewed the manuscript; Dr Panchal helped conceptualize and design the Bili-ruler, particularly the aspects related to digital color processing, and reviewed the manuscript; Dr Baqui supervised the study and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

Ms Whelan's current affiliation is Community Partners International, Yangon, Myanmar.

Dr Panchal's current affiliation is OpenBiome, Somerville, MA.

Ms Schaeffer's current affiliation is Institute for Health Metrics and Evaluation, Seattle, WA.

DOI: https://doi.org/10.1542/peds.2018-2039

Accepted for publication Feb 19, 2019

Address correspondence to Anne CC Lee, MD, MPH, Global Newborn Health Laboratory, Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, BB502A, 75 Francis St, Boston, MA 02115. E-mail: alee6@bwh.harvard.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FINANCIAL DISCLOSURE: Other than the disclosures provided in the Funding and the Potential Conflict of Interest statements, the authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: This work was supported by the US Agency for International Development's Saving Lives at Birth Program, grant AID-OAA- F-15-00023.

POTENTIAL CONFLICT OF INTEREST: Dr Lee and Ms Folger are currently employed by Brigham and Women's Hospital; Little Sparrows Technologies, Inc, has been granted a sole, royalty-bearing license by Brigham and Women's Hospital under the copyright rights to use, reproduce, and sell licensed products related to this research; the other authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Bhutani VK, Stark AR, Lazzeroni LC, et al; Initial Clinical Testing Evaluation and Risk Assessment for Universal Screening for Hyperbilirubinemia Study Group. Predischarge screening for severe neonatal hyperbilirubinemia identifies infants who need phototherapy. *J Pediatr*. 2013;162(3):477–482.e1
2. Keren R, Tremont K, Luan X, Cnaan A. Visual assessment of jaundice in term and late preterm infants. *Arch Dis Child Fetal Neonatal Ed*. 2009;94(5):F317–F322
3. 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
4. Ip S, Lau J, Chung M, et al. Hyperbilirubinemia and kernicterus: 50 years later. *Pediatrics*. 2004;114(1):263–264
5. Bhutani VK, Wong RJ. Bilirubin neurotoxicity in preterm infants: risk and prevention. *J Clin Neonatol*. 2013;2(2):61–69
6. Maisels MJ, McDonagh AF. Phototherapy for neonatal jaundice. *N Engl J Med*. 2008;358(9):920–928
7. Valaes T, Koliopoulos C, Koltsidopoulos A. The impact of phototherapy in the management of neonatal hyperbilirubinemia: comparison of historical cohorts. *Acta Paediatr*. 1996;85(3):273–276
8. Bhutani VK. Editorial: building evidence to manage newborn jaundice worldwide. *Indian J Pediatr*. 2012;79(2):253–255
9. World Health Organization. *Integrated Management of Childhood Illness Chart Booklet*. Geneva, Switzerland: World Health Organization; 2014
10. Gosset IH. A perspex icterometer for neonates. *Lancet*. 1960;1(7115):87–88
11. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527
12. Stott GJ, Lewis SM. A simple and reliable method for estimating haemoglobin. *Bull World Health Organ*. 1995;73(3):369–373
13. Lewis SM, Stott GJ, Wynn KJ. An inexpensive and reliable new haemoglobin colour scale for assessing anaemia. *J Clin Pathol*. 1998;51(1):21–24
14. Bhutani VK, Cline BK, Donaldson KM, Vreman HJ. The need to implement effective phototherapy in resource-constrained settings. *Semin Perinatol*. 2011;35(3):192–197
15. American Academy of Pediatrics Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation [published correction appears in *Pediatrics*. 2004;114(4):1138]. *Pediatrics*. 2004;114(1):297–316
16. Mantagou L, Fouzas S, Skylogianni E, Giannakopoulos I, Karatza A, Varvarigou A. Trends of transcutaneous bilirubin in neonates who develop significant hyperbilirubinemia. *Pediatrics*. 2012;130(4). Available at: www.pediatrics.org/cgi/content/full/130/4/e898
17. Taylor JA, Burgos AE, Flaherman V, et al; Better Outcomes through Research for Newborns Network. Discrepancies between transcutaneous and serum bilirubin measurements. *Pediatrics*. 2015;135(2):224–231
18. Schumacher RE, Thornbery JM, Gutcher GR. Transcutaneous bilirubinometry: a comparison of old and new methods. *Pediatrics*. 1985;76(1):10–14
19. Bilgen H, Ince Z, Ozek E, Bekiroglu N, Ors R. Transcutaneous measurement of hyperbilirubinaemia: comparison of the Minolta jaundice meter and the Ingram

icterometer. *Ann Trop Paediatr*. 1998; 18(4):325–328

20. Olusanya BO, Slusher TM, Imosemi DO, Emokpae AA. Maternal detection of neonatal jaundice during birth hospitalization using a novel two-color icterometer. *PLoS One*. 2017;12(8): e0183882
21. Gupta PC, Kumari S, Mullick DN, Lal UB. Icterometer: a useful screening tool for neonatal jaundice. *Indian Pediatr*. 1991; 28(5):473–476
22. Hamel BC. Usefulness of icterometer in black newborns with jaundice. *Trop Doct*. 1982;12(4, pt 2):213–214
23. Chaibva NT, Fenner A, Wolfsdorf J. Reliability of an icterometer in black neonates with hyperbilirubinaemia. *S Afr Med J*. 1974;48(36):1533–1534
24. Olusanya BO, Imosemi DO, Emokpae AA. Differences between transcutaneous and serum bilirubin measurements in black African neonates. *Pediatrics*. 2016;138(3):e20160907
25. United Nations Children's Fund (UNICEF). *The State of the World's Children 2017: Children in a Digital World*. New York, NY: United Nations Children's Fund (UNICEF); 2017
26. Luu MN, Le LT, Tran BH, et al. Home-use icterometry in neonatal hyperbilirubinaemia: cluster-randomised controlled trial in Vietnam. *J Paediatr Child Health*. 2014;50(9): 674–679

A Novel Icterometer for Hyperbilirubinemia Screening in Low-Resource Settings

Anne CC Lee, Lian V. Folger, Mahmoodur Rahman, Salahuddin Ahmed, Nazmun Nahar Bably, Lauren Schaeffer, Rachel Whelan, Pratik Panchal, Sayedur Rahman, Arun Dutta Roy and Abdullah H. Baqui

Pediatrics originally published online April 5, 2019;

PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

A Novel Icterometer for Hyperbilirubinemia Screening in Low-Resource Settings

Anne CC Lee, Lian V. Folger, Mahmoodur Rahman, Salahuddin Ahmed, Nazmun Nahar Bably, Lauren Schaeffer, Rachel Whelan, Pratik Panchal, Sayedur Rahman, Arun Dutta Roy and Abdullah H. Baqui

Pediatrics originally published online April 5, 2019;