

Validity of Newborn Clinical Assessment to Determine Gestational Age in Bangladesh

Anne CC Lee, MD, MPH,^{a,b} Luke C. Mullany, PhD,^b Karima Ladhani, ScM,^{a,c} Jamal Uddin, MBBS,^d Dipak Mitra, PhD,^b Parvez Ahmed, MBBS,^e Parul Christian, DrPH,^b Alain Labrique, PhD,^b Sushil K. DasGupta,^e R. Peter Lokken, MD, MPH,^f Mohammed Quaiyum, MBBS,^e Abdullah H Baqui, DrPH,^b for the Projahnmo Study Group

abstract

BACKGROUND: Gestational age (GA) is frequently unknown or inaccurate in pregnancies in low-income countries. Early identification of preterm infants may help link them to potentially life-saving interventions.

METHODS: We conducted a validation study in a community-based birth cohort in rural Bangladesh. GA was determined by pregnancy ultrasound (<20 weeks). Community health workers conducted home visits (<72 hours) to assess physical/neuromuscular signs and measure anthropometrics. The distribution, agreement, and diagnostic accuracy of different clinical methods of GA assessment were determined compared with early ultrasound dating.

RESULTS: In the live-born cohort ($n = 1066$), the mean ultrasound GA was 39.1 weeks (SD 2.0) and prevalence of preterm birth (<37 weeks) was 11.4%. Among assessed newborns ($n = 710$), the mean ultrasound GA was 39.3 weeks (SD 1.6) (8.3% preterm) and by Ballard scoring the mean GA was 38.9 weeks (SD 1.7) (12.9% preterm). The average bias of the Ballard was -0.4 weeks; however, 95% limits of agreement were wide (-4.7 to 4.0 weeks) and the accuracy for identifying preterm infants was low (sensitivity 16%, specificity 87%). Simplified methods for GA assessment had poor diagnostic accuracy for identifying preterm births (community health worker prematurity scorecard [sensitivity/specificity: 70%/27%]; Capurro [5%/96%]; Eregie [75%/58%]; Bhagwat [18%/87%], foot length <75 mm [64%/35%]; birth weight <2500 g [54%/82%]). Neonatal anthropometrics had poor to fair performance for classifying preterm infants (areas under the receiver operating curve 0.52–0.80).

CONCLUSIONS: Newborn clinical assessment of GA is challenging at the community level in low-resource settings. Anthropometrics are also inaccurate surrogate markers for GA in settings with high rates of fetal growth restriction.



^aDepartment of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, Massachusetts;

^bDepartment of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland;

^cDepartment of Global Health & Population, Harvard T. H. Chan School of Public Health, Boston, Massachusetts;

^dChild Health Research Foundation, Shishu Hospital, Dhaka, Bangladesh; ^eInternational Centre for Diarrheal Disease Research, Dhaka, Bangladesh; and ^fDepartment of Radiology, University of Illinois Hospital and Health Sciences System, Chicago, Illinois

Dr Lee conceptualized and designed the study, obtained funding, implemented the study, performed data analysis, and drafted the initial manuscript; Dr Mullany helped conceptualize and design the study, obtain funding, performed data analysis, and reviewed and revised the manuscript; Ms Ladhani performed data analysis, and reviewed and revised the manuscript; Drs Uddin, Ahmed, and Mitra helped design the data collection instruments, coordinate and supervise data collection, train physician and community health workers, and reviewed and

WHAT'S KNOWN ON THIS SUBJECT: Most preterm infants are born in and die in low-income countries where gestational age (GA) is unknown or inaccurate. Postnatal clinical assessments are sometimes used to estimate the maturity or GA of infants, primarily in high-income settings.

WHAT THIS STUDY ADDS: Compared to ultrasound dating, clinical newborn assessments of GA performed by community health workers were inaccurate, with wide margins of error (± 4 weeks) and poor diagnostic accuracy. Anthropometrics were inaccurate predictors of GA in a setting where fetal growth restriction is common.

To cite: Lee AC, Mullany LC, Ladhani K, et al. Validity of Newborn Clinical Assessment to Determine Gestational Age in Bangladesh. *Pediatrics*. 2016;138(1):e20153303

Preterm birth (<37 weeks' gestation) is the leading cause of mortality in children <5¹ and results in 1 million neonatal deaths annually.² Almost all (99%) occur in low- and middle-income countries (LMICs),¹ where preterm infants carry a seven-fold increased mortality risk compared with their full-term counterparts.³ Of the 15 million annual preterm births globally, 10 million occur in homes or first-level facilities in LMICs.⁴ In these settings, preterm infants are commonly unrecognized and/or fail to seek medical care.

Accurate and feasible methods of determining gestational age (GA) are urgently needed in LMICs to facilitate the early recognition and referral of premature infants, and the delivery of potentially life-saving interventions. Pregnancy dating is frequently uncertain in low-resource settings due to late presentation for antenatal care, challenges of last menstrual period (LMP) recall, and unavailability of ultrasonography. In high-income countries, postnatal clinical assessment of infant physical and neurologic maturity was commonly used to estimate GA before ultrasound was widely available.^{5,6} The Dubowitz and Ballard scores may predict GA \pm 14 days of LMP dating.⁶ However, these methods are complex, require neurologic examination, and computation, and, thus, may not be feasible for frontline health workers in LMICs.^{4,5} Additionally, neurologic examinations may be influenced by other morbidities, such as birth asphyxia, infection, or congenital anomalies.

Simplified methods to identify premature infants that rely on fewer characteristics,⁷ external signs only,^{8,9} or individual physical anthropometrics¹⁰⁻¹² have been described and developed for lower resource settings. The Eregie, Capurro, and Parkin scores (Supplemental Table 5) have been reported to estimate GA in high

correlation with the Dubowitz score.^{7,13} Foot length has also been explored as a potential single screening measure for prematurity and low birth weight.^{10,11}

In South Asia, another challenge is the high prevalence of fetal growth restriction, which may influence the validity of the postnatal clinical maturity assessment. Bhagwat et al¹⁴ described a simplified algorithm for GA determination (Supplemental Table 5) that correlated well with LMP-based GA in 2 hospital-based studies in India.^{14,15} Narayanan et al¹⁶ developed a 6-sign examination, including ophthalmic assessment of the anterior vascular capsule of the lens.¹⁷ When performed by physicians in a tertiary-level hospital in New Delhi, this assessment dated 95% of newborns within 11 days of LMP dates.¹⁶

The current evidence base regarding GA assessment in LMICs is limited by several factors. Clinical newborn assessments have been traditionally used by medical professionals (physicians, midwives, or nurses), and have not been evaluated when performed by nonmedically trained frontline health workers, who are often the first and only newborn contact in LMICs, or in community settings in which 40 million infants are born annually.¹⁸ Perhaps the greatest limitation is that few studies have validated GA methods against a gold standard of early ultrasound in LMICs.⁵ The aim of our study was to validate a simple prematurity scorecard as well as standard clinical assessments of GA performed by frontline community health workers (CHWs) in rural Bangladesh, as compared with early pregnancy ultrasonography.

METHODS

Study Site

This study was conducted by the Projahnmo study group¹⁹ in its

Bangladesh field site located in Sylhet district (Kanaighat and Zakiganj subdistricts: 670 km²). The Projahnmo study group is a collaboration of the Ministry of Health and Family Welfare of the Government of Bangladesh, the International Centre for Diarrheal Disease Research-Bangladesh, Shimantik nongovernmental organization, Child Health Research Foundation, Brigham and Women's Hospital/Harvard Medical School, and the Johns Hopkins Bloomberg School of Public Health. The population has an annual birth cohort of 15 000, with high baseline rates of home birth (~90%) and neonatal mortality (36.8/1000 live births).²⁰ The study area is served by CHWs: women residents of the community with at least 10th grade education, as well as 6 weeks of specialized training on basic maternal and newborn care. The CHWs for this study had on average 5 years of newborn care experience.

Pregnancy Surveillance, Eligibility, and Enrollment

This study was nested within a cluster randomized trial (clinicaltrials.gov: NCT01572532) funded by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development evaluating the impact of a community-based screening and treatment program for maternal genitourinary tract infections on the rate of preterm birth.²¹ During monthly pregnancy surveillance visits, if a period was missed, a home pregnancy test was performed, and mothers identified at <20 weeks gestation were enrolled after obtaining verbal consent. Exclusion criteria included intrauterine fetal demise, severe congenital anomalies, or withdrawal of consent. The study was approved by the Ethical Review Committees of the International Centre for Diarrheal Disease Research-Bangladesh, the Johns Hopkins Bloomberg School of

Public Health, and Partners Health Care Institutional Review Boards.

Ultrasonography

A study ultrasonographer (medical physician with ultrasound certification) was trained and standardized in early pregnancy biometry for pregnancy dating, and scans were performed in the field clinic by using a portable Nanomax Sonosite ultrasound machine (Fuji Sonosite, Inc, Bothell, WA). For fetuses <14 weeks by LMP, crown rump length was measured, and for those 14 to 19 weeks, biparietal diameter (BPD) and femoral length were also measured. Three measures of each biometric parameter were obtained. An external radiologist (PL) reviewed a random 10% of images for a quality control assessment, based on a predetermined checklist (Supplemental Figure 5). GA was estimated as per Hadlock et al, by using median crown rump length to date pregnancies <14 weeks²² and BPD for pregnancies \geq 14 weeks.²³

Neonatal Assessment

A literature review was conducted to identify existing postnatal clinical assessments and a range of potential individual neuromuscular and physical clinical signs to be included (Supplemental Table 5). Signs were performed individually during the assessment, then combined in the analytic stage into the different scoring systems. The neonatal assessment included 6 neuromuscular signs, followed by 12 physical signs and 7 anthropometrics. Signs from the Ballard, Eregie, Parkin, Capurro, and Bhagwat scores were included with minor modifications (Supplemental Table 6; Supplemental Fig 6).^{6-9,13,14,24} For the Eregie, we also tested the score by using local standards for head circumference and mid-upper arm circumference (MUAC). The assessment required 30 to 45 minutes to perform.

We also designed a simple CHW scorecard to screen for prematurity (Supplemental Fig 7). The criteria selected were most strongly correlated with GA based on previous literature, feasible for nonmedically trained providers, and culturally acceptable. The scoring system included 5 physical characteristics categorized into 3 GA categories (red zone: <34 weeks, yellow zone: 34–36 weeks, and green zone: term \geq 37 weeks). The number in each color zone was totaled, with the highest number corresponding to the assigned GA category.

Birth weight, infant length, foot length, breast bud diameter, head circumference, MUAC, and chest circumference were measured thrice. The following devices were used: KL-218 digital weighing scale (precision 10g; Dongguan Manufacturing, Hong Kong, China), JiVitA infant length board (JiVitA, Gaibandha, Bangladesh),^{25,26} and JiVitA measuring tape (JiVitA).²⁶ Foot length was measured from base of the heel to tip of the hallux with a clear plastic metric ruler (locally purchased, Sylhet, Bangladesh) using methods described by Marchant et al.¹²

A total of 24 CHWs were trained and standardized in the newborn assessment (detailed in the text of the Supplemental Information). Refresher training was conducted after 6 months.

A home visit was conducted by the CHW as soon as possible after delivery notification. Newborns visited >72 hours were excluded from the analysis. The assessment was not performed if the family refused or if the infant had signs of very severe illness. For quality control, a study physician conducted independent examinations on a random 10% of newborns, and also directly observed 5% of CHW assessments.

Data Analysis

Stata 12.0 (StataCorp, College Station, TX) was used for analyses. Preterm birth was defined as <37 weeks of gestation by early ultrasound dating. Small for gestational age (SGA) was defined as <10% birth weight for GA by using the INTERGROWTH-21st birth weight standard.²⁷ For analysis of individual signs, the correlation of scores with GA was determined by the Spearman rank correlation coefficient. The percentage of preterm births was determined for each category and the Pearson χ^2 statistic was used to determine the significance of the difference in proportions.

We assessed the agreement of gold standard ultrasound dating with postnatal GA determination by using Bland-Altman analysis to determine the mean bias (difference) and 95% limits of agreement (LOA). The Stata batplot command was used, allowing for assessment of trends and the adjustment of LOA by a regression model of the difference and averages of measures. The trend significance was tested with the Pearson correlation coefficient. Linear regression was performed to determine the trend line of mean difference. Lin's concordance analysis²⁸ was also performed to assess the correlation of GA methods.

For neonatal anthropometrics, receiver operating curves (ROCs) were generated and area under the curve (AUC) calculated for the diagnostic accuracy of anthropometrics to identify preterm births. The best anthropometric cutoff for a measure was chosen as that with the highest average sensitivity and specificity. For all methods, we calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the identification of preterm infants.

RESULTS

The pregnancy cohort was enrolled from May 2012 to December 2013 (Fig 1). A total of 1380 mothers consented, of whom 1162 were enrolled and 1066 infants were born alive. Among livebirths, mean GA was 39.1 weeks (SD 2.0) and preterm prevalence was 11.4%, with early-moderate preterm birth (<34 weeks) prevalence of 2.6%. A total of 710 newborns were assessed at <72 hours of life (651 term, 59 preterm) by a CHW. Losses to follow-up were higher in the preterm group ($n = 62$), particularly as these infants were more likely to have died ($n = 8$), been excluded for illness ($n = 14$), or born in the hospital and thus visited at >72 hours ($n = 34$) or lost to follow-up ($n = 6$). CHWs performed on average 3 to 4 newborn assessments per month, with a total of 35 assessments per CHW over the study period.

Among assessed infants, a histogram of the GA distribution is shown in Fig 2. Mean ultrasound-based GA was 39.3 weeks (SD 1.6, range 29.6–44.0), with 59 births (8.3%) <37 weeks and 7 (1.0%) <34 weeks. The mean birth weight was 2787 g (SD 416) (among term infants: 2820 g, SD 400; preterm infants: 2435 g, SD 423). The prevalence of SGA in the population was 32.4% using the INTERGROWTH-21st standard.²⁷ The average z-score for birth weight was -1.03 (SD 1.02), length -0.29 (SD 1.54), and head circumference -0.23 (SD 1.37).

Correlation of Individual Physical and Neuromuscular Signs With GA

The relationship between individual physical and neuromuscular signs and GA is shown in Tables 1 and 2. The correlation of GA with individual physical signs was low for most signs, but significant for skin texture, breast appearance, and female labia. GA was positively correlated with the individual neuromuscular signs, although the correlation coefficients were also low. Posture, scarf sign,

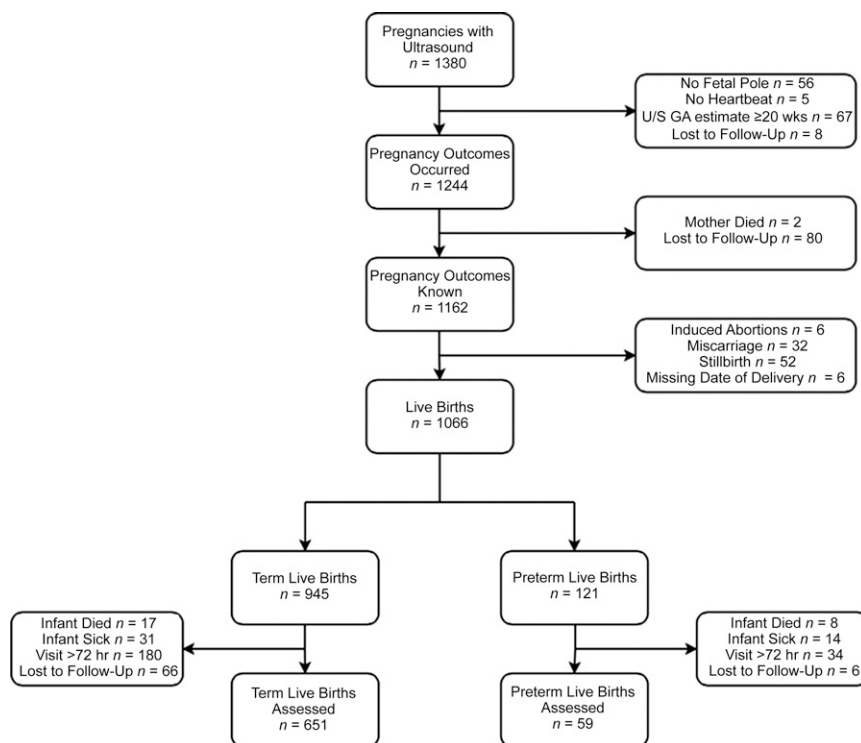


FIGURE 1
Projahnmo Saving Lives at Birth Gestational Age Validation Flowchart.

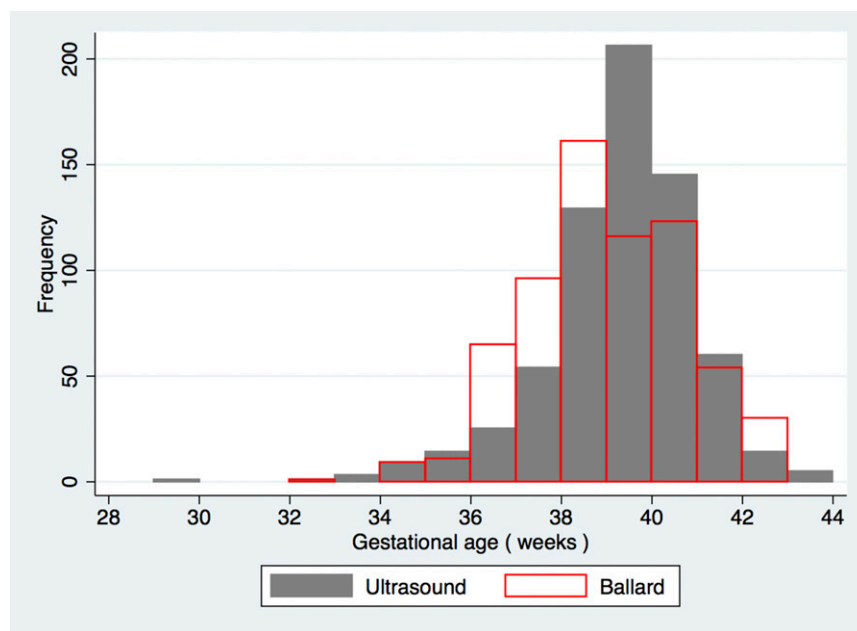


FIGURE 2
Distribution of GA by early ultrasound versus original Ballard score.

arm recoil, and ankle dorsiflexion were significantly correlated with GA. We also examined the relationship in the subset of SGA infants and

found significant correlation for skin texture and posture; however, correlation coefficients were similar to infants appropriate for GA (AGA).

TABLE 1 Correlation of Individual Physical Maturity Signs With Gestational Age

Physical Signs	Level	n	Correlation Coefficient	
			% Preterm	
Skin texture	Very thin, gelatinous, and smooth	499	9.20	0.14 (<i>< .01</i>)**
	Not thin, superficial peeling	83	6.00	
	Slight thickening, possible cracks	104	6.70	
	Thick and parchment-like, deep cracks	24	4.20	
Skin color	Dark red	4	0.00	0.05 (.17)
	Uniformly pink	534	9.20	
	Pale pink, variable color	147	5.40	
Skin opacity	Pale, only soles/palms pink	25	8.00	0.02 (.64)
	Many/several big and small veins	205	7.30	
	Few veins	204	9.80	
	Rare veins and indistinct	227	7.00	
Lanugo	No lanugo	5	0.00	−0.01 (.80)
	Abundant	208	8.20	
	Thinning, especially on back	267	6.70	
	Bald areas, little hair	169	11.20	
Ear shape	Mostly bald	57	8.80	0.02 (.57)
	Pinna flat and NO incurving	38	5.30	
	Partial incurving of whole upper pinna	82	9.80	
Ear recoil	Well-defined curving of pinna	589	8.30	0.03 (.36)
	Pinna soft and slow/easy recoil	20	5.00	
	Soft in places, ready recoil	76	7.90	
Breast appearance	Firm and thick, instant recoil	613	8.30	0.14 (<i>< .01</i>)**
	Nipple barely visible or	46	15.20*	
	Flat and smooth areola but defined	149	12.10	
	Stippled areola, not raised	514	6.40	
Male testes	Stippled and raised areola	514	6.40	0.02 (.76)
	Neither testes in scrotum	2	0.00	
	At least 1 testes low in inguinal canal	59	8.50	
Male scrotum	At least 1 testes descended	239	8.80	0.03 (.55)
	Few/faint rugae	71	8.50	
	Many/good rugae	103	9.70	
Female labia	Many deep rugae	125	8.00	0.1 (.04)*
	Majora widely separated/minora protruding	85	11.80	
	Majora almost covers minora	186	7.00	
Female clitoris	Majora completely covers minora	138	7.20	0.08 (.12)
	Prominent clitoris	111	9.00	
	Less prominent	184	8.20	
Plantar creases	Clitoris not visible	108	6.50	0.02 (.53)
	No creases/faint red marks	36	13.90	
	Anterior transverse crease only	128	6.20	
	Creases over 2/3 of anterior transverse	276	6.90	
	Creases over entire sole	252	9.10	

** *P* < .01.* *P* < .05.

Comparison of Agreement of GA Between Different Methods

In Table 3, we summarize the GA distribution of different established postnatal clinical assessment methods and report the mean bias, 95% LOA, and concordance correlation of these methods compared with ultrasound GA. Most clinical assessments had wide LOA, dating 95% of infants

within approximately ± 4 weeks of ultrasound dating.

The average GA of the cohort was similar by Ballard scoring versus ultrasound; however, the number of preterm births was higher by Ballard due to the wider distribution of GA (12.9% vs 8.3%). Among all infants, the average difference between early ultrasound and Ballard dating was -0.4 weeks (95% LOA $-4.7, 4.0$). There

was no evidence of a significant trend in the Bland-Altman plot across GA (Fig 3A). Thirty-two percent of Ballard GA estimates fell within ± 1 weeks of ultrasound dating, and 64% within ± 2 weeks. The external physical Ballard signs tended to systematically underestimate GA, whereas the neuromuscular signs slightly overestimated GA. Bland-Altman plots are shown for AGA (Fig 3B) versus SGA

TABLE 2 Correlation of Individual Neuromuscular Maturity Signs With Gestational Age

Neuromuscular Signs	Level	n	Correlation Coefficient	
			% Preterm	
Arm recoil	0–2	31	16.10	0.07
	3	163	7.40	(.05)*
	4	516	8.10	
Posture	0–2	40	25.00**	0.12
	3	164	6.70	(<.01)**
	4	506	7.50	
Popliteal angle	0–1	79	5.10*	0.05
	2	78	12.80	(.23)
	3	205	12.70	
	4	299	6.00	
	5	47	2.10	
Scarf sign	0–2	81	12.30	0.08
	3	410	8.00	(.04)*
	4	217	7.40	
	5	101	7.90	
Heel-to-ear	0–1	101	7.90	0.04
	2	191	8.90	(.26)
	3	247	10.10	
	4	170	5.30	
Ankle dorsiflexion	0–2	25	24.00**	0.08
	3	194	9.30	(.04)*
	4	490	7.10	

* $P < 0.05$.** $P < 0.01$.

infants (Fig 3C). Among SGA infants, there was evidence of a significant trend in the bias. The Ballard assessment tended to systematically underestimate GA, particularly in the lower GA ranges for SGA infants. For a 36-week SGA infant, this would equate to a 2.5-week underestimate of GA by Ballard scoring.

By Eregie examination, GA was systematically underestimated with an average bias of 2 weeks (95% LOA $-5.4, 1.5$) and left-shifting of the GA distribution (mean GA 37.5 weeks, Supplemental Fig 8A). Modification of the Eregie score to adjust head circumference and MUAC to local Bangladeshi quartiles shifted the GA distribution to a mean of 39.4 weeks; however, the modification also resulted in a narrower distribution of GA and did not classify any infants as preterm or result in improved performance. Capurro GA was biased toward overestimation by 0.4 weeks (95% LOA $-3.6, 4.5$) (Supplemental Fig 8B). The Bhagwat assessment developed in India was biased toward underestimation -0.9 (95%

TABLE 3 Agreement of Methods for Clinical Postnatal GA Determination With Early Ultrasound (GA in Weeks)

	Mean GA, wks (SD)	Median GA, wks (range)	Average Bias ^a (95% LOA)	Lin's Concordance (SE)
Early ultrasound (reference)	39.2 (1.6)	39.4 (29.6–44.0)	NA	NA
Ballard total ²⁴	38.9 (1.7)	38.8 (32.4–43.2)	-0.4 ($-4.7, 4.0$)	0.12 (0.04)
Ballard external	37.8 (2.0)	37.6 (31.2–43.2)	-1.5 ($-6.1, 3.2$)	0.10 (0.03)
Ballard neuro	39.9 (2.3)	40.0 (31.2–44.0)	0.7 ($-4.6, 6.0$)	0.08 (0.03)
Capurro ⁷	39.7 (1.6)	39.7 (32.5–43.4)	0.4 ($-3.6, 4.5$)	0.14 (0.04)
Original Eregie ¹³	37.3 (1.5)	37.5 (32.7–41.5)	-2.0 ($-5.4, 1.5$)	0.19 (0.02)
Modified Eregie ⁸	39.4 (0.6)	39.4 (38.0–40.8)	0.1 ($-2.9, 3.2$)	0.18 (0.03)
Parkin ⁹	38.6 (1.6)	38.6 (34.5–42.0)	-0.7 ($-4.8, 3.5$)	0.14 (0.04)
Bhagwat ¹⁴	38.4 (1.5)	38.5 (33.0–41.0)	-0.9 ($-5.0, 3.2$)	0.11 (0.03)

—, not applicable.

^a Average bias defined as mean difference between (Clinical GA method – early pregnancy ultrasound).

LOA $-5.0, 3.2$) (Supplemental Fig 8C). By the CHW prematurity scorecard, 240 infants (33%) were categorized in the red zone (early preterm) and 278 (39%) in the yellow zone (late-moderate preterm).

Neonatal Anthropometrics as Surrogate Markers of GA

The relationships of neonatal physical anthropometrics versus GA are shown graphically in

scatterplots in Supplemental Fig 9. The correlation coefficients ranged from 0.1 to 0.37. ROCs and AUCs for different anthropometrics as surrogate measures to identify preterm births are shown in Fig 4. The AUCs were low for foot length, infant length and MUAC (0.51–0.65), but fair for head circumference, chest circumference, and weight (0.72–0.80). The validity of the anthropometric cutoffs with the best average sensitivity/specificity are summarized in Table 4.

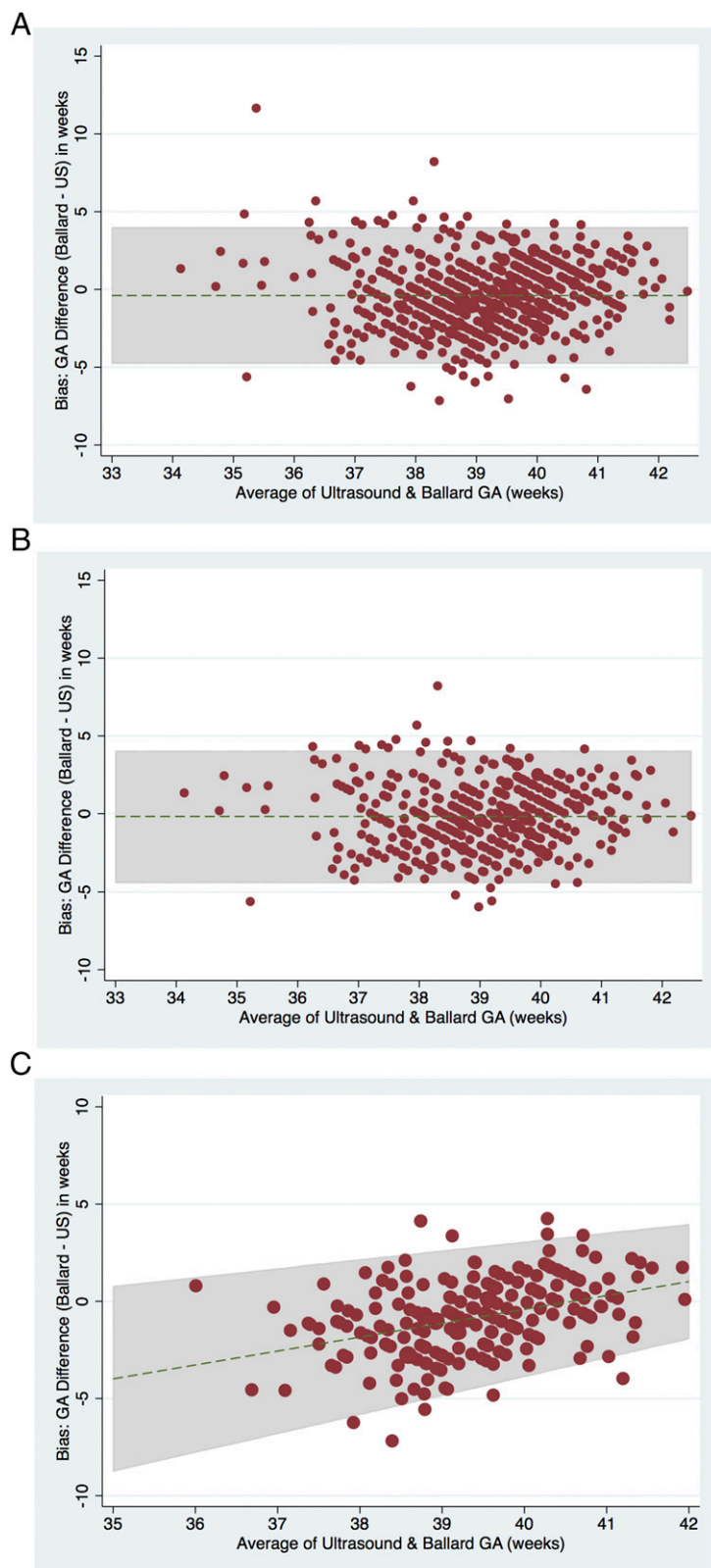


FIGURE 3
 Bland-Altman plots of Ballard versus early ultrasound for GA dating. A, All infants, no significant trend. B, AGA infants, no significant trend. C, SGA infants, significant trend line of difference ($P < .01$), bias = $0.7146235^* (\text{average Ballard_US}) - 29.00176$.

Validity of Methods for Identification of Preterm Infants

The validity of different postnatal clinical assessments tested to identify preterm infants is shown in Table 4. The Ballard, Capurro, Bhagwat, and Parkin had low sensitivity for the identification of preterm infants, although specificity was high. The Eregie and CHW prematurity scorecard had fair sensitivity (70%–75%); however, lower specificity and PPV. None of these clinical methods had adequate sensitivity or PPV to serve as a clinical screening tool in our community setting.

Surrogate neonatal anthropometrics performed slightly better; however, still did not achieve adequate sensitivity, specificity, and PPV in our setting with high rates of growth restriction. Achieving sensitivity of >70% was at the expense of specificity for all anthropometrics. Foot length was relatively nonspecific for identifying preterm births.

DISCUSSION

In our community-based Bangladeshi birth cohort with accurate early pregnancy ultrasound dating, 1 in 8 infants was born too soon (<37 weeks). This corroborates a high burden of preterm birth in a representative rural South Asian population, although the prevalence was lower than previous estimates with LMP-based dating. We validated several established and simplified postnatal methods to ascertain GA by CHWs. Standard clinical postnatal assessments, including the Ballard, Eregie, Parkin, Capurro, and Bhagwat scores, had poor validity for classifying preterm infants in our setting. The CHW prematurity scorecard had fair sensitivity but low specificity. Neonatal anthropometric measurements also had relatively poor-fair discriminatory ability for identifying preterm births where fetal growth restriction is common.

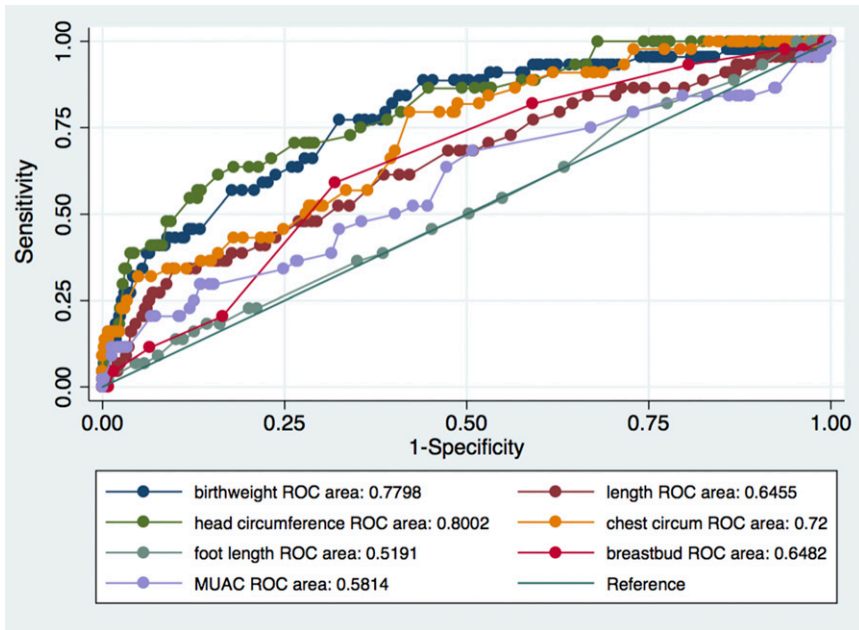


FIGURE 4 Diagnostic accuracy of physical anthropometrics to identify preterm (<37 wk) newborns.

Individual signs of physical maturity were poorly correlated with ultrasound GA in our community-based study. Previous studies have shown high correlation of most physical signs with LMP-based GA dating in mainly high-income, facility/NICU settings (correlation coefficients ranging 0.5–0.8).^{9,24} Differences in the gold standard GA determination method (ultrasound versus LMP), the low number of early preterm

infants in our study cohort, and place of assessment (home versus facility) may contribute to our findings. It is also possible that the level of health worker affected our findings. Previous validation studies have primarily used physicians; however, CHWs from our study were rigorously trained and standardized, and CHWs had high levels of agreement on individual Ballard signs compared with physicians.²⁹ In previous studies,

our CHWs have identified neonatal illness/infection with high validity compared with physicians.³⁰ Another factor potentially contributing to the performance of the physical signs is the variable time of home assessment (<72 hours of life). Certain characteristics, particularly the skin examination, may be less accurate after the first day of life. In our study, the median visit time was 13 hours and 89% of visits were within 24 hours of life.

In general, neurologic signs are more easily influenced by disease state and comorbidities, such as birth asphyxia or neonatal infections. The timing of the assessment after birth also may affect the infant’s neurologic state (ie, tone, arousability), and may have influenced our findings. Of the neurologic signs, posture, ankle dorsiflexion, arm recoil, and scarf sign scores were significantly but not strongly correlated with GA. Ankle dorsiflexion measures the relative contribution of relaxins and other parturition hormones to prepare the infant for vaginal birth (L. Dubowitz, MD, personal communication, 2012) and may be less influenced by illness.

In our community-based study, established postnatal clinical assessments had relatively wide

TABLE 4 Diagnostic Accuracy of Postnatal Clinical Methods to Identify Preterm (<37 Weeks) Infants

Method	Prevalence, %	Sensitivity, %	Specificity, %	PPV, %	NPV, %
Ballard	13	15	87	9	92
Ballard-External	35	36	65	8	92
Ballard-Neuro	14	24	86	14	93
Capurro	4	5	96	10	92
Eregie	44	75	58	14	96
Bhagwat	14	18	87	11	92
Parkin	7	10	93	12	92
CHW prematurity scorecard	72	70	27	8	91
Foot length, mm					
≤75	65	64	35	8	92
≤76	74	86	28	19	92
Birth weight, g					
≤2600	36	75	68	18	97
≤2500	21	54	82	22	95
Head circumference, cm					
≤32	20	56	83	23	95
≤33	38	68	65	15	96

NPV, negative predictive value.

LOA with ultrasound GA and poor diagnostic accuracy for identifying preterm births. The Ballard had a wide margin of error (± 4 weeks) and other simplified methods, including our CHW prematurity scorecard, had poor validity. An important point is that assessments tested in this study were designed to measure infant maturity, as opposed to gestational duration. Although prematurity is a major cause of infant immaturity, other factors may contribute, including fetal growth restriction, and maternal and neonatal morbidity. Among SGA infants, the Ballard systematically underestimated GA in lower GA ranges; thus, potentially limiting the accuracy of postnatal maturity assessment in settings with high rates of fetal growth restriction. We validated a simplified method developed in India,¹⁴ a setting with similar SGA rates; however, the diagnostic accuracy was poor. Other techniques, such as anterior vascularity of the lens, may improve performance for GA prediction in growth-restricted populations and are being studied.^{16,17,31}

The original neonatal maturity scores were developed when LMP was the gold standard, and used regression equations or scoring algorithms to predict LMP-based GA. Few studies have validated postnatal clinical assessment compared to early ultrasound GA in LMIC settings. In Malawi, Wylie et al³² reported that 79% of infants classified as preterm by the Ballard performed by research nurses were full-term according to best obstetric estimate, including ultrasound (ie, 21% PPV). Taylor and colleagues³³ found that the external Ballard performed by midwives correlated poorly with ultrasound GA in the Gambia. In Papua New Guinea, the Ballard examination performed by nurses systematically overestimated GA by 6 days with wide 95% LOA ($-27, 39$ days) and systematically underestimated GA in the lower GA ranges.³⁴

Foot length has been identified in several community-based studies as a promising indicator of preterm or low birth weight infants in Tanzania, Uganda, and Nepal.^{10,12,35} In Tanzania and Uganda, studies reported sensitivity ranging from 93% to 96% and specificity ranging from 58% to 76% for preterm birth; however, these used a gold standard diagnosis of clinical examination (Eregie/Ballard), whereas our study used a gold standard diagnosis of early pregnancy ultrasound.^{10,12} In our population in which fetal growth restriction is prevalent, foot length was not an accurate predictor of prematurity.

The main limitation of this study is the survival bias of those infants who were included in the analysis. Preterm, particularly early preterm, infants were more commonly excluded, either because they died before the visit, were delivered in a hospital, or were too ill for the assessment. Early preterm infants more likely would have been identified as preterm by the test screening methods. Thus, we believe our estimates would tend to underestimate the sensitivity to identify preterm births in the entire birth cohort. Our study environment, however, would reflect real-life performance of these methods in a community setting with high rates of home birth and home visitation. In these settings, earlier contact with preterm infants, by improved prebirth pregnancy dating, notification of preterm labor, and earlier home visits, may more effectively improve referral and outcomes. Furthermore, the clinical assessment may be more beneficial and have improved performance in health facilities, where staff are more skilled and frequent assessment may improve performance. In facility settings in which most infants may be assessed at birth, the early identification of

preterm infants may lead to earlier delivery of interventions to improve outcomes. Finally, our analysis reflects the diagnostic accuracy of these methods and anthropometrics in a South Asian setting and may not be generalized to other settings, particularly African settings in which growth restriction is less common.

This work is the result of a pilot study. An expanded assessment is being tested in a multicenter study as part of the World Health Organization (WHO) Alliance for Maternal and Newborn Health Improvement (AMANHI) study. The AMANHI study will validate a comprehensive neonatal assessment, including feeding assessment, in a larger sample size of 7000 newborns in 5 countries (Bangladesh, Ghana, Pakistan, Pemba-Tanzania, and Zambia). This study will have a wider range of signs tested, increased statistical power, and will include facility-based births and assessments by higher-level health workers. It will also use advanced computational machine learning analysis to test multiple permutations, combinations, and simplified algorithms for identifying preterm births and will enable comparisons and generalizations to different world regions.

CONCLUSIONS

In our community-level validation study in Bangladesh, assessment of postnatal clinical maturity by CHWs had poor diagnostic accuracy to identify preterm infants as defined by early pregnancy ultrasound. Neonatal anthropometrics were also relatively nonspecific to identify preterm births in our setting with high rates of fetal growth restriction. The delay to first newborn contact for home births, particularly among preterm infants, is a major barrier to

improve preterm birth outcomes in LMICs. There is an urgent need to improve pregnancy dating before birth, reduce delays to first newborn contact, and develop methods to feasibly and accurately identify preterm births to improve birth outcomes in settings of highest need.

ACKNOWLEDGMENTS

We thank all the staff of the Projahnmo team with special thanks to those who implemented this project, including Rina Paul, Eusuf Ashraf, Monir Zaman, Mahmood Rahman, Aatur Rahim, Tasima Hussain, Nasreen Islam, field supervisors, project officers, CHWs, and Village Health Workers. We

thank our partners at WHO and the AMANHI principal investigators and teams: WHO: Rajiv Bahl and Alex Manu, Pakistan: Fyezah Jehan, Pemba: Sunil Sazawal, Zambia: David Hamer, Ghana: Betty Kirkwood, Lisa Hurt, Massachusetts General Hospital: Blair Wylie. We thank Lily Dubowitz, Jeanne Ballard, Kris Karlsen (STABLE program), and Jeanne Aby for providing their inputs into the assessment and for sharing training photographs, video, and/or pictograms for trainings. We thank Rachel Whelan, Lian Folger, and Chelsea Clarke for their assistance in formatting and preparing the manuscript for submission. Finally, we extend our appreciation to the families, mothers, and infants who participated in this study.

ABBREVIATIONS

AGA: appropriate for gestational age
AMANHI: Alliance for Maternal and Newborn Health Improvement
AUC: area under the curve
BPD: biparietal diameter
CHW: community health worker
GA: gestational age
LMIC: low-and middle-income country
LMP: last menstrual period
LOA: limits of agreement
MUAC: mid-upper arm circumference
PPV: positive predictive value
ROC: receiver operating curve
SGA: small for gestational age
WHO: World Health Organization

revised the manuscript; Dr Christian helped conceptualize the design of the study protocols particularly related to pregnancy ultrasonography, provided input on data analysis, and reviewed and revised the manuscript; Drs Labrique and Quaiyum helped conceptualize the design of the study, and reviewed and revised the manuscript; Mr DasGupta helped design the data collection instruments, data management system, and reviewed and revised the manuscript; Dr Lokken helped conceptualize the study procedures for ultrasonography, reviewed and provided quality control measures for ultrasound measures, and reviewed and revised the manuscript; Dr Baqui helped conceptualize and design the study, obtain funding, provided input on data analysis, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

This trial has been registered at www.clinicaltrials.gov (identifier NCT01572532).

DOI: 10.1542/peds.2015-3303

Accepted for publication Mar 30, 2016

Address correspondence to Anne CC Lee, MD, MPH, Brigham and Women's Department of Pediatric Newborn Medicine, 75 Francis St, Thorn 229a, Boston, MA 02115. E-mail: alee6@partners.org

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

Copyright © 2016 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: This study is made possible through the generous support of the Saving Lives at Birth Round 1 partners: the US Agency for International Development, the Government of Norway, the Bill & Melinda Gates Foundation, the World Bank, and Grand Challenges Canada. It was prepared by the Projahnmo research group and does not necessarily reflect the views of the Saving Lives at Birth Partners. The study was also funded by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (R01 HD066156-02). Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

COMPANION PAPER: A companion to this article can be found online at www.pediatrics.org/cgi/doi/10.1542/peds.2016-0734.

REFERENCES

1. Liu L, Johnson HL, Cousens S, et al; Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 2012;379(9832):2151–2161
2. Oza S, Lawn JE, Hogan DR, Mathers C, Cousens SN. Neonatal cause-of-death estimates for the early and late neonatal periods for 194 countries: 2000–2013. *Bull World Health Organ*. 2015;93(1):19–28
3. Katz J, Lee AC, Kozuki N, et al; CHERG Small-for-Gestational-Age-Preterm Birth Working Group. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *Lancet*. 2013;382(9890):417–425
4. World Health Organization, The Partnership for Maternal, Newborn, and Child Health, Save the Children; March of Dimes. *Born Too Soon: the Global Action Report on Preterm Birth*.

- Geneva, Switzerland: World Health Organization; 2012
5. Dubowitz LM, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. *J Pediatr.* 1970;77(1):1–10
 6. Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard Score, expanded to include extremely premature infants. *J Pediatr.* 1991;119(3):417–423
 7. Capurro H, Konichezky S, Fonseca D, Caldeyro-Barcia R. A simplified method for diagnosis of gestational age in the newborn infant. *J Pediatr.* 1978;93(1):120–122
 8. Eregie CO. A new method for maturity determination in newborn infants. *J Trop Pediatr.* 2000;46(3):140–144
 9. Parkin JM, Hey EN, Clowes JS. Rapid assessment of gestational age at birth. *Arch Dis Child.* 1976;51(4):259–263
 10. Nabweya E, Marchant T, Namazzi G, Kadobera D, Waiswa P. Identifying high-risk babies born in the community using foot length measurement at birth in Uganda. *Child Care Health Dev.* 2013;39(1):20–26
 11. Mukherjee S, Roy P, Mitra S, Samanta M, Chatterjee S. Measuring newborn foot length to identify small babies in need of extra care: a cross-sectional hospital based study. *Iran J Pediatr.* 2013;23(5):508–512
 12. Marchant T, Jaribu J, Penfold S, Tanner M, Armstrong Schellenberg J. Measuring newborn foot length to identify small babies in need of extra care: a cross sectional hospital based study with community follow-up in Tanzania. *BMC Public Health.* 2010;10:624
 13. Eregie CO, Muogbo DC. A simplified method of estimating gestational age in an African population. *Dev Med Child Neurol.* 1991;33(2):146–152
 14. Bhagwat VA, Dahat HB, Bapat NG. Determination of gestational age of newborns—a comparative study. *Indian Pediatr.* 1990;27(3):272–275
 15. Bindusha S, Rasalam CS, Sreedevi N. Gestational age assessment of newborn- clinical trial of a simplified method. *Transworld Medical Journal.* 2014;2(1):24–28
 16. Narayanan I, Dua K, Gujral VV, Mehta DK, Mathew M, Prabhakar AK. A simple method of assessment of gestational age in newborn infants. *Pediatrics.* 1982;69(1):27–32
 17. Hittner HM, Hirsch NJ, Rudolph AJ. Assessment of gestational age by examination of the anterior vascular capsule of the lens. *J Pediatr.* 1977;91(3):455–458
 18. Lawn JE, Blencowe H, Oza S, et al; Lancet Every Newborn Study Group. Every newborn: progress, priorities, and potential beyond survival. *Lancet.* 2014;384(9938):189–205
 19. Baqui AH, El-Arifeen S, Darmstadt GL, et al; Projahnmo Study Group. Effect of community-based newborn-care intervention package implemented through two service-delivery strategies in Sylhet district, Bangladesh: a cluster-randomised controlled trial. *Lancet.* 2008;371(9628):1936–1944
 20. Arifeen SE, Mullany LC, Shah R, et al. The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial. *Lancet.* 2012;379(9820):1022–1028
 21. Lee AC, Quaiyum MA, Mullany LC, et al. Screening and treatment of maternal genitourinary tract infections in early pregnancy to prevent preterm birth in rural Sylhet, Bangladesh: a cluster randomized trial. *BMC Pregnancy Childbirth.* 2015;15:326
 22. Hadlock FP, Shah YP, Kanon DJ, Lindsey JV. Fetal crown-rump length: reevaluation of relation to menstrual age (5–18 weeks) with high-resolution real-time US. *Radiology.* 1992;182(2):501–505
 23. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal biparietal diameter: a critical re-evaluation of the relation to menstrual age by means of real-time ultrasound. *J Ultrasound Med.* 1982;1(3):97–104
 24. Ballard JL, Novak KK, Driver M. A simplified score for assessment of fetal maturation of newly born infants. *J Pediatr.* 1979;95(5 pt 1):769–774
 25. Labrique AB, Shaikh S, West KP, inventors; Alain Labrique, assignee. Portable acrylic infant length board. US patent 61/450,949/2011. February 2011
 26. Labrique AB, Christian P, Klemm RD, et al. A cluster-randomized, placebo-controlled, maternal vitamin A or beta-carotene supplementation trial in Bangladesh: design and methods. *Trials.* 2011;12:102
 27. Villar J, Cheikh Ismail L, Victora CG, et al; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet.* 2014;384(9946):857–868
 28. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989;45(1):255–268
 29. Lee Anne CC, Uddin J, Shah RML, et al. Validation of community health worker clinical assessment of gestational age in rural Bangladesh. In: Proceeding from the Pediatric Academic Societies Annual Meeting; May 4 - May 7, 2013; Washington, DC
 30. Baqui AHAS, Arifeen SE, Rosen HE, et al; Projahnmo Study Group. Community-based validation of assessment of newborn illnesses by trained community health workers in Sylhet district of Bangladesh. *Trop Med Int Health.* 2009;14(12):1448–1456
 31. Skapinker R, Rothberg AD. Postnatal regression of the tunica vasculosa lentis. *J Perinatol.* 1987;7(4):279–281
 32. Wylie BJ, Kalilani-Phiri L, Madanitsa M, et al. Gestational age assessment in malaria pregnancy cohorts: a prospective ultrasound demonstration project in Malawi. *Malar J.* 2013;12:183
 33. Taylor RADF, Denison FC, Beyai S, Owens S. The external Ballard examination does not accurately assess the gestational age of infants born at home in a rural community of The Gambia. *Ann Trop Paediatr.* 2010;30(3):197–204
 34. Karl S, Li Wai Suen CSN, Unger HW, et al. Preterm or not—an evaluation of estimates of gestational age in a cohort of women from rural Papua New Guinea. *PLoS One.* 2015;10(5):e0124286
 35. Mullany LC, Darmstadt GL, Khatri SK, Leclercq SC, Tielsch JM. Relationship between the surrogate anthropometric measures, foot length and chest circumference and birth weight among newborns of Sarlahi, Nepal. *Eur J Clin Nutr.* 2007;61(1):40–46