Correcting for measurement error in assessing gestational age in a low-resource setting: a regression calibration approach

George O. Agogo[1]*, Jennifer R. Verani[1], Nancy A. Otieno[2],
Bryan O. Nyawanda[2], Marc-Alain Widdowson[1,3] and
Sandra S. Chaves[4]

[1]Division of Global Health Protection, US Centers for Disease Control and Prevention, Nairobi, Kenya,
[2]Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya, [3]Institute of
Tropical Medicine, Antwerp, Belgium, [4]Influenza Program, US Centers for Disease Control and
Prevention, Nairobi, Kenya

Introduction: Measurement error in gestational age (GA) may bias the association of GA with a health outcome. Ultrasound-based GA is considered the gold standard and is not readily available in low-resource settings. We corrected for measurement error in GA based on fundal height (FH) and date of last menstrual period (LMP) using ultrasound from the sub-cohort and adjusted for the bias in associating GA with neonatal mortality and low birth weight (< 2,500 grams, LBW).

Methods: We used data collected from 01/2015 to 09/2019 from pregnant women enrolled at two public hospitals in Siaya county, Kenya ($N = 2,750$). We used regression calibration to correct for measurement error in FH- and LMP-based GA accounting for maternal and child characteristics. We applied logistic regression to associate GA with neonatal mortality and low birth weight, with and without calibrating FH- and LMP-based GA.

Results: Calibration improved the precision of LMP (correlation coefficient, ρ from 0.48 to 0.57) and FH-based GA (ρ from 0.82 to 0.83). Calibrating FH/LMP-based GA eliminated the bias in the mean GA estimates. The log odds ratio that quantifies the association of GA with neonatal mortality increased by 29% (from −0.159 to −0.205) by calibrating FH-based GA and by more than twofold (from −0.158 to −0.471) by calibrating LMP-based GA.

Conclusion: Calibrating FH/LMP-based GA improved the accuracy and precision of GA estimates and strengthened the association of GA with neonatal mortality/ LBW. When assessing GA, neonatal public health and clinical interventions may benefit from calibration modeling in settings where ultrasound may not be fully available.

KEYWORDS

attenuation, fundal height, gestational age, last menstrual period, measurement error, preterm, regression calibration

Introduction

Gestational age (GA) at birth is closely linked with neonatal and infant health outcomes. Accurate assessment of GA is important to guide patient care and public health interventions to improve pregnancy outcomes and infant health (1). More than 60% of deaths in the first year of life are either associated with preterm birth (< 37 weeks of gestation) or with low birth weight (< 2,500 grams) (1, 2). In extremely preterm births, causes of death include respiratory disorders and failure. Mortality rates among preterm infants correlate with birth weight and GA, with low birth weight and preterm births associated with high burden of morbidity and poorer survival. Therefore, infants born with the lowest GA and birth weight have the greatest risk of death (2). Fetal ultrasound is widely used in high-income settings as the gold standard, and biometric measurements provide the most accurate prediction of expected date of delivery, especially if performed before 20 weeks of gestation (3, 4). In low-resource settings, access to ultrasound equipment and skilled technicians is limited, particularly in rural and remote areas (5). In addition, pregnant women in low- and middle-income countries (LMIC) often do not seek prenatal care until later in pregnancy, which further limits the use of ultrasound to assess GA (4, 6). Kenya recommends that a pregnant woman should attend a minimum of four comprehensive personalized antenatal visits spread out during the entire pregnancy as follows: first (< 16 weeks), second (16 to <28 weeks), third (28 to <32 weeks), and fourth (≥ 32 weeks) visit (7). In low-resource settings, such as Kenya, GA is commonly assessed using more affordable but error-prone methods such as the last menstrual period (LMP) or fundal height (FH) (8, 9). Dating based on LMP can be unreliable due to imperfect maternal recall and variable timing of fertilization relative to ovulation (1, 8, 10). Similarly, the accuracy of FH may be affected by multiple pregnancies, maternal size, intrauterine growth restriction, and fetal position (3). Inaccurate GA measurements using these error-prone methods can over- or underestimate the burden of preterm births, which can, in turn, impact clinical care, health services planning, and health policy decisions (11). Moreover, the measurement error in LMP and FH will bias associations of GA with health outcomes and reduce statistical power to detect significant associations (10).

Studies have focused mainly on assessing the validity and accuracy of alternative GA assessment methods relative to the ultrasound (12, 13). Currently, there is limited focus on adjusting for the measurement error in GA before associating GA with a health outcome. We demonstrated a statistical method that corrects for the bias in the parameter estimate that quantifies the association of GA with the health outcome. To demonstrate the method, we used data from a cohort study of pregnant women in rural western Kenya, where GA was measured by ultrasound, FH, and LMP. We used the ultrasound data available from a subset of the women as the gold standard to correct for measurement error in FH- and LMP-based GA in the main study and generated measurement error–corrected GA estimates for all study participants. Subsequently, we used the measurement error–corrected GA values to associate GA with birth outcomes. We also compared FH with LMP regarding their level of precision in measuring GA.

Materials and methods

Study population

From January 2015 to September 2019, the Kenya Medical Research Institute (KEMRI) and the Centers for Disease Control and Prevention (CDC) in Kenya conducted a cohort study of pregnant women to assess the burden of influenza disease and its impact on birth outcomes in two rural public hospitals in Siaya County, namely, Siaya County Referral Hospital and Bondo sub-County Hospital (14). The study area has a high burden of maternal and infant mortality, HIV, and malaria, as described in Otieno et al. (14). Pregnancy was confirmed by a blood test or ultrasound, and maternal HIV status was assessed at enrolment, as required by the Kenyan Ministry of Health guidelines, and prophylaxis was provided when indicated. The participants were either recruited during home visits or routine antenatal care visits at the study hospitals. The inclusion criteria included consenting pregnant women at ≤30 weeks gestation, aged 15 to 49 years, resident of a village within 10 kilometers of the study health facility, consenting to HIV counselling and testing, not planning to relocate out of the study area, not enrolled in another interventional study, agreeing to all follow-up visits, and willingness to deliver in the labor ward of the study hospitals (15). The exclusion criteria for the study included multiple pregnancies (twins, triplets, or high-order multiples), history of fistula repair or leg/spinal deformity, or those unable to give informed consent. Participants were followed weekly through either phone calls or home visits until delivery. Structured questionnaires were used to collect demographic and clinical data, including obstetric history, prior pregnancies, medical history, physical examination, and baseline laboratory tests.

Gestational age estimation

Gestational age was assessed at enrolment by ultrasound, FH, and/or LMP. Ultrasound machines were available initially through routine care as part of the Ministry of Health antenatal care follow-up. However, they were not consistently available to the study staff due to competing priority use for routine care in other services, technician unavailability, or malfunctioning of the machine. In 2018, an ultrasound machine (DP-10 Mindray) was purchased for each study site, and clinicians were trained before using the ultrasound to measure GA. Using this machine, ultrasound GA was calculated according to the parameters obtained in the following measurements: GA in obstetrics (OB) items, average ultrasound age (AUA), and composite ultrasound age (CUA). The GA in the OB items was calculated by the related GA formulae, and AUA is the average of GAs that are calculated according to biparietal diameter, head circumference, abdomen circumference, femur length, gestational sac, and crown rump length, among other parameters. The CUA is calculated using formulae based on biparietal diameter, head circumference, abdomen circumference, and femur length (16). Further details on using this ultrasound machine to measure GA can be found in the operator's manual (16). Despite that, the ultrasound results were available only for a subset of the study participants. In measuring GA using FH, women were instructed to empty their bladder, a non-elastic tape was used with the graduation in centimeters, and the number of centimeters was considered to correspond to GA in weeks (17). A fundal height was taken as the distance measured from the highest point of the uterus (fundus) to the top of the symphysis pubis. In measuring GA using LMP, the study participants were asked

about the date of their last LMP. Last menstrual period was defined as the first day of the last menstruation (4).

Regression calibration

We used a regression calibration method to correct for measurement error in FH- and LMP-based GA before relating GA to a health outcome. Using regression calibration, calibrated GA was defined as the conditional expectation of true GA given error-prone FH- or LMP-based GA and predictors of GA assumed to be measured without error (hereafter, error-free predictors) (18, 19). The calibration model was developed from a subset of data derived from the participants with observed ultrasound, LMP, FH, and error-free predictors; then, the resulting GA estimates from the calibration model were used to obtain calibrated GA for all the participants with and without ultrasound measurements. The calibrated GA estimates were then used to associate GA with health outcomes instead of the error-prone measurements from FH/LMP. Using this method, we corrected for measurement error in FH and LMP separately and then jointly. In Figure 1, we denote measurement error in LMP and FH by δ_{lmp} and δ_{fh}, respectively, a true parameter that quantifies the association of GA with a health outcome by λ_T and the biased association parameter by λ_Q. The regression calibration method adjusts for the bias in estimating the association parameter λ_T. Using regression calibration, the estimate of the unknown true GA (denoted by T) is a function of GA based on LMP and/or FH (denoted by Q) and error-free predictors (denoted by Z) (18, 20–22). We denoted regression calibration by E (T|Q, Z). Thus, the unknown true association parameter, λ_T was estimated by using the calibrated GA or the error-prone GA measurements from LMP/FH in the outcome regression analysis (23). Ultrasound-based GA measurement in the calibration model was assumed to be unbiased for

true GA and to be measured with random error (δ_{US}) that is uncorrelated with true GA and measurement error in LMP and FH (18, 19, 24). To calibrate GA based on FH using regression calibration, ultrasound measurements were regressed on FH measurements and a set of error-free predictors of GA. Calibrated GA based on FH was obtained as the conditional mean GA estimates from the calibration model for those with and without ultrasound, provided all other variables were measured and included in the calibration model. To calibrate GA based on LMP and FH jointly, ultrasound-based GA was regressed on LMP- and FH-based GA and a set of error-free predictors.

To adjust for the bias in estimating the association parameter λ_T, we used the calibrated GA to associate GA with neonatal mortality (deaths within the first 28 days of life, encoded as dead/alive) and low birth weight (< 2,500 grams, yes/no). We modeled the two binary outcomes separately using logistic regression and quantified the association with the logarithm of the odds ratio (hereafter, log OR). To account for the uncertainty in estimating the association parameter due to the calibration process, the standard errors for the calibrated log OR estimates were obtained using the bootstrap method (25–27). With the bootstrap method, the standard error was calculated as the standard deviation of the log OR estimates from the bootstrapped samples. To assess the effect of calibrating GA, we further estimated uncalibrated log OR using GA based on FH and LMP in the logistic regression to associate GA with each outcome.

Statistical analysis

Gestational age from LMP, FH, and ultrasound was summarized using descriptive statistics. We fitted a linear regression calibration model using ultrasound-based GA measurements as the response under the normality assumption. To eliminate implausible values in
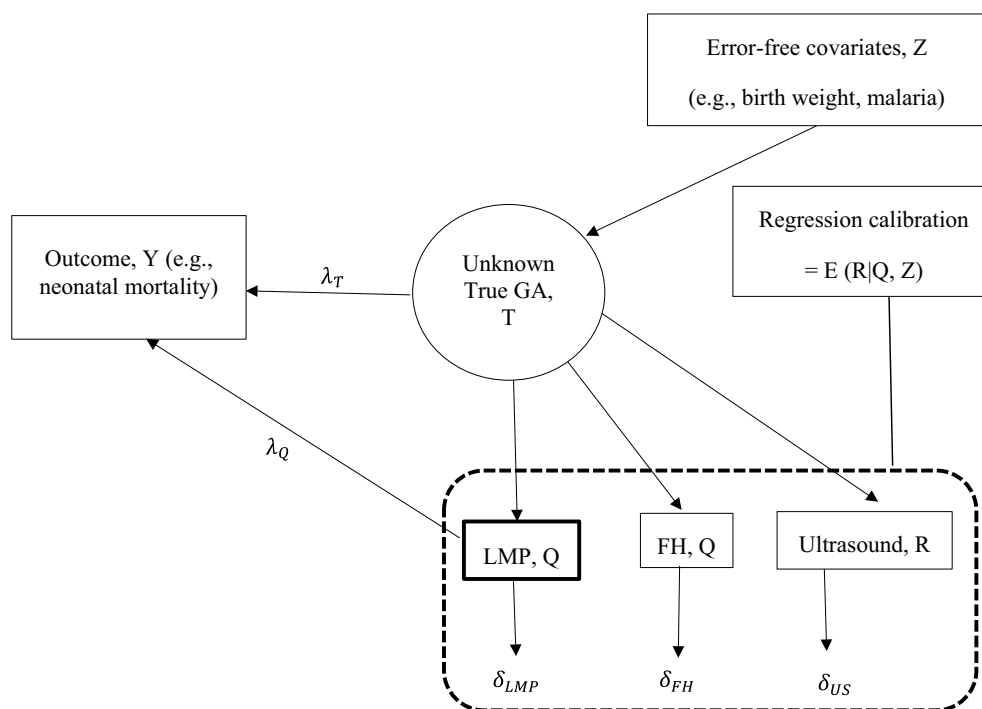

FIGURE 1
Gestational age (GA) assessment, measurement error, and association of GA with outcome.

the response, we excluded ultrasound measurements ≤0.5th (i.e., 28 weeks gestation) or ≥ 99.5th (i.e., 44 weeks gestation) percentile of the distribution of ultrasound values; these implausible ultrasound values were considered extreme. We assumed the following variables as error-free predictors and included them in the regression calibration model: baby's birth weight and sex, and for mothers, their age, HIV status, mid-upper arm circumference, malaria and syphilis diagnoses, first pregnancy, education level, trimester of current pregnancy, and hemoglobin levels at enrolment and delivery. The selection of these predictors was guided by their associations with GA at birth from previous studies (28–30). We, however, excluded the baby's birth weight from the list of error-free predictors when investigating the association of GA with low birth weight. To explore calibrated GA for those with and without ultrasound, we compared their corresponding distributions based on kernel/normal densities and histograms. To quantify the validity of calibrated GA relative to ultrasound-based GA, we used Pearson correlation coefficient, with a value close to 1 implying high validity and a value close to 0 implying low validity. We further used the Bland–Altman method to quantify the agreement between the calibrated GA estimates and ultrasound (31, 32). Using the Bland–Altman method, the discrepancy in the mean GA measurements was quantified by computing the mean of the difference between the calibrated GA and the GA based on ultrasound. A mean difference of 0 indicates no discrepancy in the mean GA values and can be construed as unbiased. The Bland–Altman method further quantifies the precision by constructing 95% limits of agreement. The 95% limits of agreement were calculated as the mean difference ± 2 standard deviations of the mean difference (31, 32). A narrow limit of agreement corresponds to more precise measurements (33). To present the discrepancy in the mean GA measurements graphically, we plotted the difference in GA from the two methods against their mean.

Subsequently, we estimated the association parameter using calibrated GA and hypothesized that the calibration process strengthens the association. To obtain calibrated log OR estimates, we used calibrated GA in the fitted logistic regression analyses that associate GA with neonatal mortality and low birth weight and compared the calibrated log OR estimates with uncalibrated log OR estimates. To obtain the standard error of the log OR estimate, we performed simple random sampling to obtain 1,000 bootstrap samples each of size 1,500. In each bootstrap sample, the log OR was estimated and the standard error obtained as the standard deviation of the log OR estimates from 1,000 bootstrap samples. To assess the effect of ultrasound measurements availability on the performance of regression calibration, we restricted the analysis to pregnant women with all the three GA measurements from ultrasound, FH, and LMP. Subsequently, we created random subsets based on percent ultrasound availability (25, 50 and 100%) by generating a random indicator variable, r, using a Bernoulli random number generator, where ultrasound values are set to missing if r is zero. Statistical analyses were conducted in SAS software version 9.4 (SAS Institute Inc., Cary, NC). The level of significance was defined as $\alpha \le 0.05$.

Results

Study data description

From January 2015 to September 2019, data for 2,750 pregnant women who enrolled in the cohort study were analyzed (Table 1). The median age was 24.5 years (range: 13.3 to 49.7), and 25.8% of the women

TABLE 1 Descriptive characteristics of study participants from a cohort of pregnant mothers enrolled from January 2015 to September 2019 in Siaya, Western Kenya, **N** = 2,750.

Study variables	Summaries
Maternal age (years)	
Mean (SD)	25.4 (5.5)
Median (range)	24.5 (13.3–49.7)
Maternal education – *n* (%)	
None	27 (0.98)
Primary	1,537 (55.90)
Secondary	850 (30.90)
Tertiary	336 (12.20)
First pregnancy – *n* (%)	
Yes	708 (25.8)
No	2042 (74.2)
Trimester at enrollment (estimated)[a] – *n* (%)	
First	749 (27.2)
Second	1988 (72.3)
Third	13 (0.5)
Malaria – *n* (%)	
Yes	440 (16.0)
No	2,310 (84.0)
Syphilis – *n* (%)	
Yes	24 (0.87)
No	2,726 (99.13)
Maternal HIV status – *n* (%)	
Positive	513 (18.7)
Negative	2,237 (81.3)
Hemoglobin at enrolment (g/dL)	
Mean (SD)	11.3 (1.7)
Median (range)	11.4 (4.3–19.2)
Hemoglobin at delivery (g/dL)	
Mean (SD)	10.2 (2.3)
Median (range)	10.0 (5.1–18.0)
Maternal mid-upper arm circumference (MUAC) in centimeters	
Mean (SD)	27.3 (3.0)
Median (range)	27.0 (8–45)
Baby's sex– *n* (%)	
Female	1,341 (48.8)
Male	1,409 (51.2)
Baby's birth weight (grams)	
Mean (SD)	3,209 (506)
Median (range)	3,210 (1100–5,000)

[a]Trimester of gestation was estimated by study team at enrollment based on ultrasound, date of last menstrual period, and/or fundal height obtained at enrollment (without calibration).

were in their first pregnancy. At enrollment, 749 (27.2%) participants were estimated to be in their first trimester of pregnancy, 1,988 (72.3%) in their second trimester, and 13 (0.5%) in their third trimester. Slightly more than half (55.9%) of the participants had attained a primary level of education,

and 16% were diagnosed with malaria and 19% with HIV. Among newborns, 51.2% were male, the mean (standard deviation, SD) weight was 3209 (506) grams, and the median weight was 3,210 grams. There were 29 (1.1%) deaths during the neonatal period, and 145 (5.3%) newborns were considered to have low birth weight.

Gestational age estimates

Ultrasound-based GA data were available for 1,176 (42.8%) pregnant women, FH for 2,521 (91.7%), and LMP for 2,720 (98.9%). The three assessment methods were used together in 1,068 women, constituting 38.8% of all the enrolled women. In fitting the calibration models, we excluded 19 records with potentially extreme ultrasound values. The observed means for GA estimates were 38.7 weeks using the LMP-based method, 38.8 weeks using FH, and 38.6 weeks using the calibrated LMP/FH-based method (Supplementary Table S1). The FH-based mean (SD) GA was 38.8 (2.92), and the calibrated FH-based GA estimate was 38.6 (2.05) weeks gestation. Similarly, the LMP-based mean (SD) GA was 38.7 (3.36) weeks, and the calibrated LMP-based GA was 38.6 (1.27) weeks gestation. The calibrated GA estimates showed less variability (smaller SD) than the uncalibrated GA estimates. The GA obtained by jointly calibrating FH and LMP correlated best with ultrasound (Pearson correlation, $\rho = 0.84$), followed by calibrated FH alone ($\rho = 0.83$), and then calibrated LMP alone ($\rho = 0.57$). Calibration improved the correlation between GA based on FH and ultrasound slightly from 0.82 to 0.83 and between LMP and ultrasound from 0.48 to 0.57. Calibrating GA based on FH and/or LMP eliminated the discrepancy in the mean GA estimates relative to the ultrasound-based mean GA (mean difference = 0, Table 2; Figure 2) and improved the precision of the GA estimates (tighter 95% limits of agreement for calibrated GA estimates). Based on the 95% limits of agreement, GA estimated using calibrated FH, on average, may be 2.39 weeks lower/higher than the GA estimated using ultrasound. Similarly, GA estimated using calibrated LMP-based GA on average, may be 3.56 weeks lower/higher than the GA estimated using ultrasound. Thus, the calibrated FH-based GA was more precise than the calibrated LMP-based GA, as shown by the tighter 95% limits of agreement (Table 2). The most precise GA estimates were obtained by calibrating FH- and LMP-based GA jointly.

We observed a similar mean distribution of calibrated GA between those with and without ultrasound measurements from the histograms and density plots (Figure 3). However, due to the large sample size, we observed a statistically significant difference in the mean distribution of calibrated FH among those with and without ultrasound (p value = 0.001) and of calibrated LMP among those with and without ultrasound (p value = 0.009). The percentage of births that would be classified as preterm birth changed from 24.3 to 26.5% by calibrating the FH-based GA and from 19.9 to 11.3% by calibrating the LMP-based GA, suggesting more misclassification when using LMP than FH.

Association estimates of GA with birth outcomes

In estimating the association of GA with neonatal mortality and low birth weight using continuous GA measurements, the log OR estimates from the calibrated GA were larger, in absolute values, than the log OR estimates from the uncalibrated GA, implying a strengthened association (Table 3). An increase in GA resulted in a significantly lower odds of neonatal deaths regardless of the calibration status of GA measurements. By comparing the calibrated and uncalibrated association estimates for GA with neonatal mortality, the strength of association increased by about 29% (log OR from −0.159 to −0.205) by calibrating GA based on FH, translating into a 3.8% (OR from 0.853 to 0.815) further reduction in the odds ratio of neonatal death. Similarly, we observed an 8.3% (OR from 0.703 to 0.620) further reduction in the odds ratio of neonatal death by calibrating GA based on LMP. We observed a similar result for the low-birth-weight outcome. The smaller log OR estimates (or OR estimate close to one) from the uncalibrated FH- and LMP-based GA implied the attenuation effect of measurement error in GA. The standard error estimates for the log OR that was adjusted for measurement error using regression calibration were larger than the unadjusted estimates due to the uncertainty in calibrating GA. This is further shown by the wider 95% confidence interval (CI) for the calibrated OR estimate. We observed significantly higher odds of neonatal death among preterm births relative to full-term births by calibrating GA based on FH. The odds of

TABLE 2 Descriptive measures for estimated gestational age (in weeks) at delivery from a cohort of pregnant mothers with paired GA measurements enrolled from January 2015 to September 2019 in Siaya, Western Kenya.

GA assessment methods	N	Mean	SD	Median	25th–75th percentile	Minimum maximum	Mean difference[a] (SD)	p-value Ho: bias = 0	95% limits of agreement
Ultrasound	1,176	38.7	2.15	39.0	38.0–40.0	29.0–43.0			
Fundal height (FH)									
Uncalibrated	1,097	39.0	2.45	39.0	38.0–41.0	30.0–47.0	0.25 (1.40)	<0.0001	−2.56–3.05
Calibrated	1,089	38.7	1.79	38.8	37.7–40.0	32.1–44.4	0.00 (1.22)	1.000	−2.39–2.39
Last menstrual period (LMP)									
Uncalibrated	1,154	38.7	3.36	39.0	37.0–40.0	21.0–56.0	−0.01 (3.00)	0.811	−6.00–5.98
Calibrated LMP	1,089	38.7	1.79	38.8	37.7–40.0	32.1–44.4	0.00 (1.22)	1.000	−3.56–3.56
LMP and FH									
Calibrated	1,069	38.7	1.82	38.8	37.6–39.9	32.0–43.9	0.00 (1.19)	1.000	−2.36–2.36

[a]Mean difference refers to the discrepancy between mean GA estimates from FH/LMP and the mean estimate from the ultrasound; mean estimate from ultrasound is assumed to be unbiased for the true mean GA.
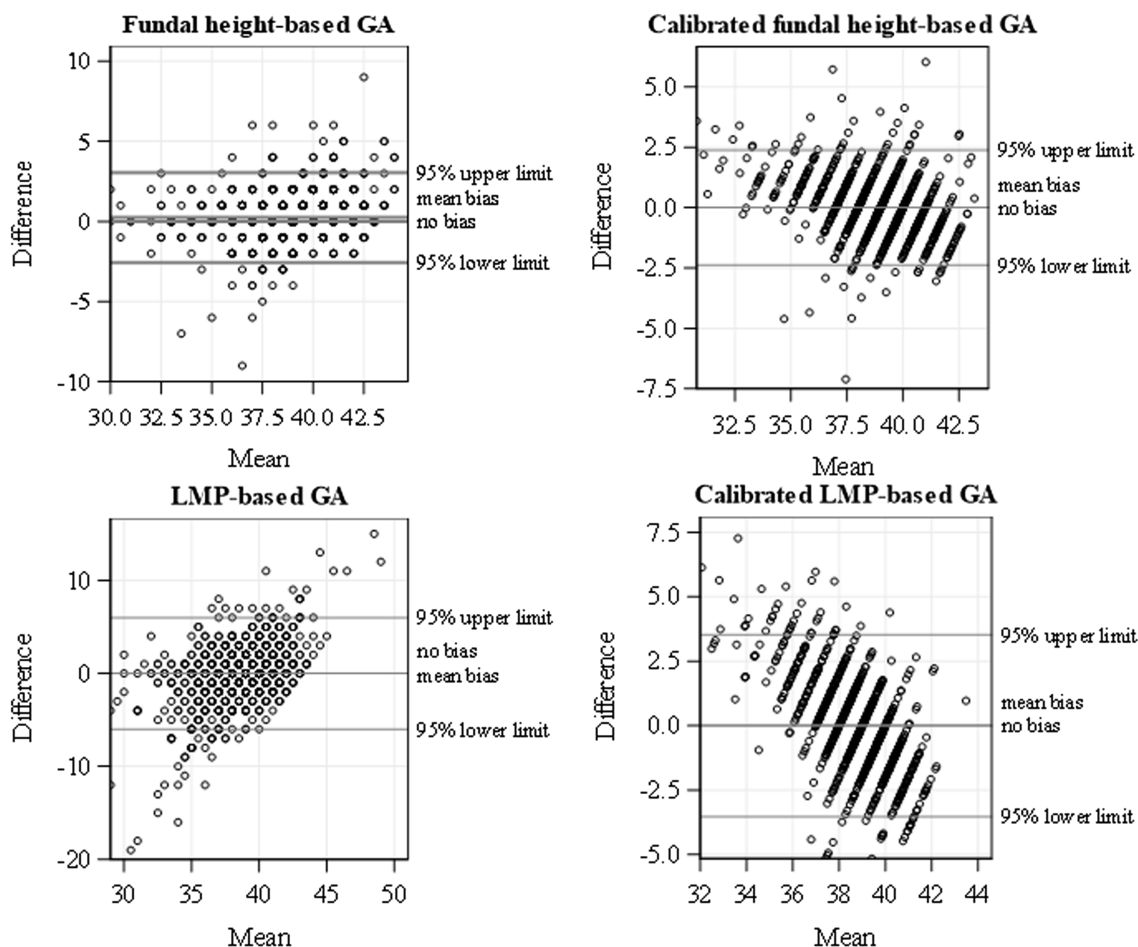
FIGURE 2
Bland–Altman plot for uncalibrated and calibrated FH and LMP showing the scatterplots of difference in GA and mean with ultrasound and the 95% limits of agreement using data from a cohort of pregnant women conducted from January 2015 to September 2019 in Siaya, Western Kenya.

a baby having low birth weight were significantly higher among preterm births relative to full-term births regardless of the GA estimation method used. From the sensitivity analysis, the proportion of available ultrasound measurements impacted the calibrated log OR estimates and standard errors, and the impact differed when one measure (FH or LMP) was calibrated compared to when both measures (FH and LMP) were calibrated (Supplementary Table S2). As a general remark, it is advisable to have as many ultrasound measurements as possible to obtain more precise effect estimates.

Discussion

Accurate assessment of GA is important for obstetric and pediatric care and guides the timing of interventions at birth in reducing neonatal morbidity and mortality (5). Regarding this, current studies mainly focus on assessing the accuracy of GA assessment tools or on the predicted GA (5, 8). In this study, we demonstrated a statistical method that corrects for measurement error in FH- and LMP-based GA before relating GA to a health outcome. The proposed regression calibration approach uses ultrasound data available in a sub-cohort to correct for measurement

error in FH- and LMP-based GA in the main cohort. Calibration improved the correlation between FH/LMP-based GA estimates and ultrasound and eliminated the mean discrepancy in GA estimates relative to ultrasound. Calibrating FH- and LMP-based GA strengthened the association of GA with neonatal mortality and low birth weight, resulting in more valid study conclusions (11, 18, 34). Of note, the calibration parameters may be used to adjust for measurement error in FH and/or LMP in a similar cohort study, provided the parameter estimates from the calibration sub-study are transferable to the cohort to be calibrated (10). However, calibration leads to association estimates with wider CIs due to the uncertainty involved in calibrating GA. Thus, regression calibration does not correct for the loss of statistical power to detect significant associations. Calibration improved the precision of GA estimates, which could affect the estimated prevalence of preterm births as shown in this study, especially when using LMP to measure GA. Low-resource settings may have limited access to ultrasound services, and our analysis provides an alternative way to improve GA estimates when ultrasound results are partially available. The observed higher correlation between FH and ultrasound compared to between LMP and ultrasound is in line with the literature (3, 4), suggesting that FH may be a more precise measure of GA than LMP
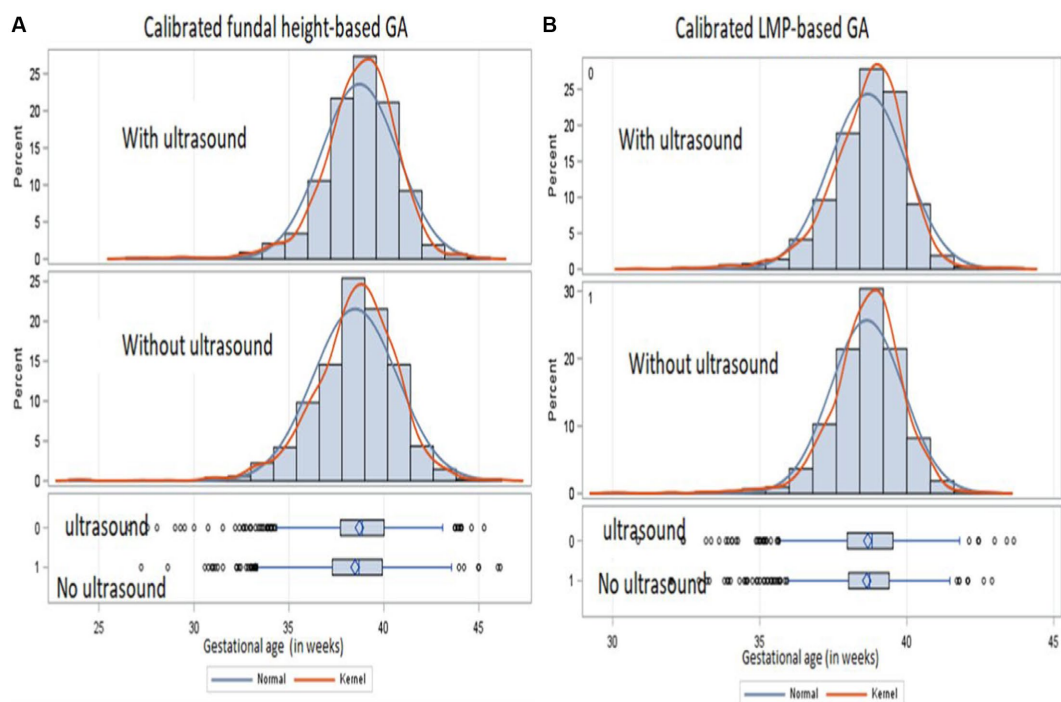
FIGURE 3
Distribution of calibrated GA from FH **(A)** and calibrated LMP **(B)** for those with ultrasound (upper panel) and those without ultrasound (lower panel) from a cohort of pregnant mothers conducted from January 2015 to September 2019 in Siaya, Western Kenya.

TABLE 3 Log odd ratio estimates and standard error for the association of gestational age with neonatal mortality and low birth weight using calibrated and uncalibrated fundal height and last menstrual period measurements.

	Association of gestational age with neonatal mortality [N = 29 (1.05%)]				Association of gestational age with low birth weight [low birth weight[b], N = 145 (5.27%)]			
	Continuous version		Categorized version (preterm vs. full term)		Continuous version		Categorized version (preterm vs. full term)	
GA assessment method	Log OR (SE)	OR (95% CI)	Log OR (SE)	OR (95% CI)	Log OR (SE)	OR (95% CI)	Log OR (SE)	OR (95% CI)
Fundal height								
Uncalibrated	−0.159 (0.050)	0.853 (0.773;0.897)	0.197 (0.204)	1.218 (0.816;1.816)	−0.353 (0.029)	0.703 (0.664;0.744)	0.669 (0.088)	1.952 (1.643;2.320)
Calibrated	−0.205 (0.090[a])	0.815 (0.683;0.891)	0.217 (0.321[a])	1.242 (0.662;2.331)	−0.478 (0.041)	0.620 (0.572;0.672)	0.662 (0.102[a])	1.939 (1.587;2.368)
Last menstrual period								
Uncalibrated	−0.158 (0.040)	0.854 (0.789;0.889)	0.327 (0.204)	1.387 (0.930;2.069)	−0.273 (0.022)	0.761 (0.729;0.795)	0.860 (0.089)	2.363 (1.985;2.814)
Calibrated	−0.471 (0.097[a])	0.624 (0.516;0.688)	0.580 (0.216[a])	1.786 (1.170;2.727)	−0.839 (0.049)	0.432 (0.393;0.476)	0.830 (0.121[a])	2.293 (1.809;2.907)

[a]Standard error for the calibrated estimate was calculated using the bootstrap method.
[b]Low birth weight defined as a birth born with less than 2,500 grams.

and could be a better method in routine antenatal care and clinical management (4, 35). Accurate assessment of GA requires proper training of healthcare workers on GA assessment (33). Compared to women with accurate GA to inform the estimated delivery date, women with moderate or severe underestimation of GA may be more likely to miss delivery at the health facility. Notably, measurement error in GA can lead to substantial error in the

estimated fetal weight, resulting in misclassification of small-for-gestational-age (SGA) and large-for-gestational-age (LGA) neonates. Small-for-gestational-age neonates are at increased risk of stillbirth and adverse perinatal outcome. Similarly, LGA neonates are at increased risk of perinatal death, birth injury, and adverse neonatal outcomes (36). Measurement error in GA can alter clinical decisions, with either an appropriate-for-gestational-age (AGA) fetus being

misclassified as SGA or LGA, an SGA or LGA fetus being misclassified as AGA, or a preterm birth being misclassified as full term. An underestimation of GA can lead to a delay in the induction of pregnancies that have entered the post-term period, increasing the risk of perinatal and neonatal mortality (37). Furthermore, misclassification of GA can impact the decision of whether to administer corticosteroids for lung maturation before the anticipated preterm birth (37). Therefore, there is a need for finding simple but accurate solutions to improve GA estimation in low-resource settings to improve clinical management in providing antenatal care (38). Subsequently, a regression calibration approach can be integrated within the antenatal care system to obtain more accurate individualized GA estimates. This can help clinicians with better classification of fetal weight and preterm births so that neonates can receive appropriate and timely clinical management and medical care (38). Moreover, antenatal care attendance may be limited in LMICs, and some women may not have opportunities for multiple antenatal visits before delivery, as was the case in our cohort study, where only one measurement was available (39).

Although the regression calibration method has been applied in other epidemiologic studies, especially in nutritional epidemiology (18, 20), this study is an innovative application of the method in improving GA estimation. Regression calibration is a relatively simple technique and can be implemented in standard software such as Stata, R, and SAS. In this study, we applied the calibration method to correct for measurement error in error-prone but more affordable GA assessment based on LMP and FH using ultrasound data available only for a subgroup of the study cohort. This approach is important in settings where it may be impossible to measure GA with ultrasound for all pregnant mothers and where multiple measurements throughout gestation may not be feasible due to costs and logistics. For accurate GA estimates from the calibration model, we recommend that ultrasound for estimating GA be conducted at <20 weeks of gestation, at which the biologic variations in fetal size and the effects of growth restriction are still small, and that the calibration model be specified correctly.

Nonetheless, our study has a few limitations. Despite the training of technicians, the reliability of the ultrasound results could have been hampered by inadequate experience, technical expertise of the technicians, and the stage of the pregnancy. In our study, GA was assessed only once, at enrollment, using different methods. Therefore, we were unable to quantify the magnitude of attenuation bias due to measurement error in FH -and LMP-based GA; quantifying attenuation requires multiple measurements per person from each assessment method (5). Additionally, the study participants were from the same ethnicity and geographical region, limiting generalizability to populations with greater demographic diversity or different background rates of health events, e.g., malaria and HIV. Lastly, in applying regression calibration, the additional covariates in the model, such as baby birth weight and trimester, were assumed to be perfectly measured, which may not hold.

Conclusion

Accurate assessment of GA in low-resource settings poses a challenge to obstetric and pediatric care management. When assessing GA, neonatal public health and clinical interventions may benefit from a regression calibration approach in settings where ultrasound may not be fully available.

Data availability statement

The datasets presented in this article are available from the corresponding author, GA, ovl6@cdc.gov, after permission from the relevant authority to release the data upon reasonable request.

Ethics statement

Ethical clearance for the study was obtained from KEMRI Scientific and Ethics Review Unit (KEMRI SSC. 2880) and CDC's institutional review board (CDC Protocol #6709) in February 2015. Informed consent was obtained from all participants before enrolment in the study. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

GA and SC conceptualized the study objective and the data analytic methods. GA analyzed the data and drafted the first manuscript. BN and NO cleaned the data and wrote the study design section. GA, JV, NO, BN, M-AW, and SC were involved in writing and revising the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1222772/full#supplementary-material

References

1. Wier ML, Pearl M, Kharrazi M. Gestational age estimation on United States livebirth certificates: a historical overview. *Paediatr Perinat Epidemiol*. (2007) 21:4–12. doi: 10.1111/j.1365-3016.2007.00856.x

2. MacDorman MF, Minino AM, Strobino DM, Guyer B. Annual summary of vital statistics—2001. *Pediatrics*. (2002) 110:1037–52. doi: 10.1542/peds.110.6.1037

3. Jehan I, Zaidi S, Rizvi S, Mobeen N, McClure EM, Munoz B, et al. Dating gestational age by last menstrual period, symphysis-fundal height, and ultrasound in Urban Pakistan. *Int J Gynaecol Obstet*. (2010) 110:231–4. doi: 10.1016/j.ijgo.2010.03.030

4. Unger H, Thriemer K, Ley B, Tinto H, Traoré M, Valea I, et al. The assessment of gestational age: a comparison of different methods from a malaria pregnancy cohort in sub-Saharan Africa. *BMC Pregnancy Childbirth*. (2019) 19:1–9. doi: 10.1186/s12884-018-2128-z

5. Pugh SJ, Ortega-Villa AM, Grobman W, Newman RB, Owen J, Wing DA, et al. Estimating gestational age at birth from fundal height and additional anthropometrics: a prospective cohort study. *BJOG*. (2018) 125:1397–404. doi: 10.1111/1471-0528.15179

6. Unger HW, Ome-Kaius M, Wangnapi RA, Umbers AJ, Hanieh S, Suen CSLW, et al. Sulphadoxine-pyrimethamine plus azithromycin for the prevention of low birthweight in Papua New Guinea: a randomised controlled trial. *BMC Med*. (2015) 13:9. doi: 10.1186/s12916-014-0258-3

7. Health Mo. (2014). *National Guidelines for quality obstetrics and perinatal care Ministry of Public Health and Sanitation and Ministry of Medical Services*. Available at: http://guidelines.health.go.ke:8000/media/National_Guidelines_for_Quality_Obstetrics_and_Perinatal_Care.pdf (Accessed January 19, 2022).

8. Rosenberg RE, Ahmed ANU, Ahmed S, Saha SK, Chowdhury MA, Black RE, et al. Determining gestational age in a low-resource setting: validity of last menstrual period. *J Health Popul Nutr*. (2009) 27:332. doi: 10.3329/jhpn.v27i3.3375

9. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol*. (2007) 21:86–96. doi: 10.1111/j.1365-3016.2007.00865.x

10. Alexander GR, Tompkins ME, Petersen DJ, Hulsey TC, Mor J. Discordance between Lmp-based and clinically estimated gestational age: implications for research, programs, and policy. *Public Health Rep*. (1995) 110:395–402.

11. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement error in nonlinear models: a modern perspective*. New York, NY: Chapman and hall/CRC (2006).

12. de Sousa Basso NA, Morceli G, Costa R, Dias A, Rudge MVC, IMP C. Validation of a symphysis-fundal height chart developed for pregnancy complicated by diabetes and hyperglycemia: an observational study. *Reprod Health*. (2016) 13:89. doi: 10.1186/s12978-016-0202-4

13. Curti A, Zanello M, De Maggio I, Moro E, Simonazzi G, Rizzo N, et al. Multivariable evaluation of term birth weight: a comparison between ultrasound biometry and symphysis-fundal height. *J Matern Fetal Neonatal Med*. (2014) 27:1328–32. doi: 10.3109/14767058.2013.858241

14. Otieno NA, Nyawanda BO, McMorrow M, Oneko M, Omollo D, Lidechi S, et al. The burden of influenza among Kenyan pregnant and postpartum women and their infants, 2015–2020. *Influenza Other Respir Viruses*. (2022) 16:452–61. doi: 10.1111/irv.12950

15. Odhiambo FO, Laserson KF, Sewe M, Hamel MJ, Feikin DR, Adazu K, et al. Profile: the KEMRI/CDC health and demographic surveillance system—Western Kenya. *Int J Epidemiol*. (2012) 41:977–87. doi: 10.1093/ije/dys108

16. Mindray. (2014). *Digital ultrasonic diagnostic imaging system Opoerator's manual*. Shenzhen Mindray bio-medical electronics co., ltd. Available at: https://www.mindraynorthamerica.com/cmsAdmin/uploads/h-046-003961-00-4-0_dc-n3-advanced-ops-manual.pdf (Accessed January 19, 2022).

17. Euans DW, Connor PD, Hahn RG, Rodney WM, Arheart KL. A comparison of manual and ultrasound measurements of fundal height. *J Fam Pract*. (1995) 40:233–6.

18. Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am J Epidemiol*. (1999) 150:642–51. doi: 10.1093/oxfordjournals.aje.a010063

19. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, et al. Structure of dietary measurement error: results of the open biomarker study. *Am J Epidemiol*. (2003) 158:14–21. doi: 10.1093/aje/kwg091

20. Agogo GO, van der Voet H, Van't Veer P, Ferrari P, Leenders M, Muller DC, et al. Use of two-part regression calibration model to correct for measurement error in episodically consumed foods in a single-replicate study design: epic case study. *PLoS One*. (2014) 9:e113160. doi: 10.1371/journal.pone.0113160

21. Spiegelman D, McDermott A, B R. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*. (1997) 65:1179S–86S. doi: 10.1093/ajcn/65.4.1179S

22. IR W. Commentary: dealing with measurement error: multiple imputation or regression calibration? *Int J Epidemiol*. (2006) 35:1081–2. doi: 10.1093/ije/dyl139

23. Rosner B, RJAjoe G. Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am J Epidemiol*. (2001) 154:827–35. doi: 10.1093/aje/154.9.827

24. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst*. (2011) 103:1086–92. doi: 10.1093/Jnci/Djr189

25. Agogo GO. A zero-augmented generalized gamma regression calibration to adjust for covariate measurement error: a case of an episodically consumed dietary intake. *Biom J*. (2017) 59:94–109. doi: 10.1002/bimj.201600043

26. Agogo GO, van der Voet H, Van't Veer P, van Eeuwijk FA, Boshuizen HC. Evaluation of a two-part regression calibration to adjust for dietary exposure measurement error in the cox proportional hazards model: a simulation study. *Biom J*. (2016) 58:766–82. doi: 10.1002/bimj.201500009

27. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. (2006) 25:4279–92. doi: 10.1002/sim.2673

28. Berkowitz GS, Blackmore-Prince C, Lapinski RH, Savitz DA. Risk factors for preterm birth subtypes. *Epidemiology*. (1998) 9:279–85. doi: 10.1097/00001648-199805000-00011

29. Lang JM, Lieberman E, Cohen A. A comparison of risk factors for preterm labor and term small-for-gestational-age birth. *Epidemiology*. (1996) 7:369–76. doi: 10.1097/00001648-199607000-00006

30. Hoffman HJ, LS B. Risk factors associated with the occurrence of preterm birth. *Clin Obstet Gynecol*. (1984) 27:539–52. doi: 10.1097/00003081-198409000-00004

31. D G. Understanding bland altman analysis. *Biochem Med (Zagreb)*. (2015) 25:141–51. doi: 10.11613/BM.2015.015

32. Bland JM, DJ A. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. (1986) 327:307–10. doi: 10.1016/S0140-6736(86)90837-8

33. Karl S, Suen CSLW, Unger HW, Ome-Kaius M, Mola G, White L, 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:e0124286. doi: 10.1371/journal.pone.0124286

34. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol*. (2012) 175:340–7. doi: 10.1093/Aje/Kwr317

35. White LJ, Lee SJ, Stepniewska K, Simpson JA, Dwell SLM, Arunjerdja R, et al. Estimation of gestational age from fundal height: a solution for resource-poor settings. *J R Soc Interface*. (2011) 9:503–10. doi: 10.1098/rsif.2011.0376

36. Wright D, Wright A, Smith E, Nicolaides KH. Impact of biometric measurement error on identification of small- and large-for-gestational-age fetuses. *Ultrasound Obstet Gynecol*. (2020) 55:170–6. doi: 10.1002/uog.21909

37. Skalkidou A, Kullinger M, Georgakis MK, Kieler H, Kesmodel US. Systematic misclassification of gestational age by ultrasound biometry: implications for clinical practice and research methodology in the Nordic countries. *Acta Obstet Gynecol Scand*. (2018) 97:440:444. doi: 10.1111/aogs.13300

38. Fulcher I, Hedt K, Marealle S, Tibaijuka J, Abdalla O, Hofmann R, et al. Errors in estimated gestational ages reduce the likelihood of health facility deliveries: results from an observational cohort study in Zanzibar. *BMC Health Serv Res*. (2020) 20:50. doi: 10.1186/s12913-020-4904-5

39. Jiwani SS, Amouzou-Aguirre A, Carvajal L, Chou D, Keita Y, Moran AC, et al. Timing and number of antenatal care contacts in low and middle-income countries: analysis in the countdown to 2030 priority countries. *J Glob Health*. (2020) 10:010502. doi: 10.7189/jogh.10.010502