

From **Every Newborn-INDEPTH study: Improving the measurement of pregnancy outcomes in population-based surveys**

RESEARCH

Open Access



Gestational age data completeness, quality and validity in population-based surveys: EN-INDEPTH study

M. Moinuddin Haider^{1*}, Kaiser Mahmud¹, Hannah Blencowe², Tahmeed Ahmed³, Joseph Akuze^{2,4,5}, Simon Cousens^{2,6}, Nafisa Delwar¹, Ane B. Fisker^{7,8,9}, Victoria Ponce Hardy², S. M. Tafsir Hasan³, Md. Ali Imam¹, Dan Kajungu¹⁰, Md Alfazal Khan¹¹, Justiniano S. D. Martins⁷, Quamrun Nahar¹, Obed Ernest A. Nettey¹², Adane Kebede Tesega^{13,14}, Judith Yargawa², Nurul Alam^{1†}, Joy E. Lawn^{2†} and the Every Newborn-INDEPTH Study Collaborative Group

Abstract

Background: Preterm birth (gestational age (GA) <37 weeks) is the leading cause of child mortality worldwide. However, GA is rarely assessed in population-based surveys, the major data source in low/middle-income countries. We examined the performance of new questions to measure GA in household surveys, a subset of which had linked early pregnancy ultrasound GA data.

Methods: The EN-INDEPTH population-based survey of 69,176 women was undertaken (2017-2018) in five Health and Demographic Surveillance System sites in Bangladesh, Ethiopia, Ghana, Guinea-Bissau and Uganda. We included questions regarding GA in months (GA_m) for all women and GA in weeks (GA_w) for a subset; we also asked if the baby was 'born before expected' to estimate preterm birth rates. Survey data were linked to surveillance data in two sites, and to ultrasound pregnancy dating at <24 weeks in one site. We assessed completeness and quality of reported GA. We examined the validity of estimated preterm birth rates by sensitivity and specificity, over/under-reporting of GA_w in survey compared to ultrasound by multinomial logistic regression, and explored perceptions about GA and barriers and enablers to its reporting using focus group discussions ($n = 29$).

Results: GA_m questions were almost universally answered, but heaping on 9 months resulted in underestimation of preterm birth rates. Preference for reporting GA_w in even numbers was evident, resulting in heaping at 36 weeks; hence, over-estimating preterm birth rates, except in Matlab where the peak was at 38 weeks. Questions regarding 'born before expected' were answered but gave implausibly low preterm birth rates in most sites. Applying ultrasound as the gold standard in Matlab site, sensitivity of survey-GA_w for detecting preterm birth (GA_w <37) was 60% and specificity was 93%. Focus group findings suggest that women perceive GA to be important, but usually counted in months. Antenatal care attendance, women's education and health cards may improve reporting.

(Continued on next page)

* Correspondence: moin@icddr.org

†Nurul Alam and Joy E Lawn are joint senior authors.

¹Health Systems and Population Studies Division, icddr,b, Dhaka, Bangladesh

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: This is the first published study assessing GA reporting in surveys, compared with the gold standard of ultrasound. Reporting GAw within 5 years' recall is feasible with high completeness, but accuracy is affected by heaping. Compared to ultrasound-GAw, results are reasonably specific, but sensitivity needs to be improved. We propose revised questions based on the study findings for further testing and validation in settings where pregnancy ultrasound data and/or last menstrual period dates/GA recorded in pregnancy are available. Specific training of interviewers is recommended.

Keywords: Gestational age, Preterm birth, Household survey, Demography, Surveillance, Last menstrual period, Ultrasound

Key findings

WHAT IS NEW?

- **What was known already:** Gestational age (GA) is instrumental in ascertaining foetal maturity and identifying preterm births; however, it is rarely assessed in population-based surveys. Quality of survey GA data, and barriers and enablers to GA data collection in such surveys have been unstudied.
- **What was done:** Analyses of a population-based survey of 69,176 women of reproductive age including novel questions on GA to assess feasibility and quality (completeness, heaping), as well as acceptability (qualitative data) in five HDSS sites, plus validity against gold standard by early ultrasound in one site (Matlab, Bangladesh).

WHAT WAS FOUND IN THE QUANTITATIVE DATA?

- **Completeness:** GA was reported for almost all births in all sites. Data on GAw was more variable. In four sites, interviewers prompted women leading to an estimate of GAw for 56–98% of births. In Bandim (Guinea-Bissau), where no prompting was used, only 6% were able to report GAw.
- **Data quality (heaping):** In Matlab (Bangladesh), survey-reported GA in months and weeks yielded similar preterm birth rates. In the other four sites, reported GAw heaped at 9 months, underestimating preterm birth rate and GAw heaped at even numbers, particularly 36 weeks, overestimating preterm birth rate.
- **Validity:** Compared to early pregnancy ultrasound, in Matlab ($n = 481$), the sensitivity of survey GAw was 60% with specificity of 93%. The sensitivity of HDSS-GAw, where date of last menstrual period was recorded in early pregnancy with an early pregnancy test was 66% and specificity was 95%.

WHAT WAS FOUND IN THE QUALITATIVE DATA?

- **Perceived value:** Women know the importance of tracking GA, notably for birth planning. Women count GA in months, not in weeks. Counting GAw from missed periods is common practice facilitated by religious and cultural events, crop harvesting times etc.
- **Barriers/enablers:** Barriers to reporting GA include lack of awareness of menstrual cycles, not retaining health cards and fear of social stigma and witchcraft.

WHAT NEXT IN MEASUREMENT AND RESEARCH?

- **Measurement improvement now:** Whilst heaping may remain a challenge, we note that other variables such as birth weight are collected in surveys despite considerable heaping and missing data. More investment and innovation are warranted given the importance of GA data for estimating preterm birth rates and data gaps in the highest-burden settings. Based on the findings in this study, we propose a revised set of questions to collect GAw.
- **Research needed:** Further studies to refine GA collection methods, link to card data and improve consistency in probing could lead to more robust approaches to assess GA in surveys. Innovation with dating apps and improving women's awareness of menstrual cycle dating are also key.

Background

Preterm birth is the leading cause of child deaths worldwide, causing an estimated one million deaths per year and a high burden of morbidity for children and their families [1–3]. Each year, an estimated 15 million babies are born preterm, the majority (91%, 13.6 million) in low- and middle-income countries (LMICs) with over 80% in Asia and sub-Saharan Africa. Accurate and timely data on preterm birth are needed to inform appropriate resources and interventions and to monitor trends. The World Health Organization (WHO) has committed to providing updated estimates of preterm births every 3 to 5 years to support progress towards targets such as the Sustainable Development Goals and the Every Newborn Action Plan, aiming to end preventable neonatal deaths and stillbirths by 2030 [1, 2]. However, substantial gaps remain in the data, especially from the highest-burden settings.

WHO defines preterm birth as any birth before 37 completed weeks of gestation as measured from the 1st day of the last menstrual period (LMP) (Table 1) [4, 5]. Measurement of gestational age (GA) is essential for identifying preterm births [10, 11]. The 'gold standard' measure of GA is to assess the baby's crown-rump length by ultrasound during early pregnancy (<14 weeks). Accuracy of ultrasound scan before 24 weeks is also considered acceptable since the difference in ultrasound-GAs measured between ≤ 13 weeks and $14 \leq 23$ weeks is less than 1 week and falls within 95% confidence interval [12, 13]. Ultrasound measures at later gestations are less accurate [14]. However, in countries with the highest burden of preterm births, the timing of the first antenatal care (ANC) visit is typically in the second trimester and access to ultrasound is limited [15]. Hence, GA is commonly assessed from the date of the last menstrual period (LMP) [10]. This method has the advantage that it can be measured at any point during pregnancy, but accuracy is highest when recorded early in pregnancy [16]. LMP has lower accuracy (± 2 –3 weeks) when compared to early pregnancy ultrasound scans [16–23]. Additionally, lower socio-

Table 1 Overview of definitions and measurements relating to gestational age

Gestational age	Measure of pregnancy duration in weeks, from the first day of the woman's last menstrual cycle to the date of assessment
Preterm birth	Preterm birth is defined by WHO as 'any birth before 37 completed weeks of gestation, or fewer than 259 days since the first day of the women's last menstrual period (LMP)' [4, 5]. It is subdivided into extremely preterm (< 28 weeks); very preterm (28-< 32 weeks) and moderate or late preterm (32-< 37 weeks) [6].
Post-term birth	Post-term birth is defined by WHO as 'any birth at/after 42 completed weeks of gestation since the first day of the women's last menstrual period (LMP)' [4, 5].
Measuring gestational age before birth	<p><i>Ultrasound (gold standard)</i>: Accurate to ± 5 to ± 21 days depending on the timing of ultrasound at different stage of pregnancy and biometric parameters (crown-rump length, biparietal diameter, femur length, abdominal circumference, head circumference).</p> <p>CRL was used for pregnancy of < 14 weeks, and BPD and FL were used for ≥ 14 to < 24 weeks.</p> <p><i>Last menstrual period (LMP)</i>: Women's recall of the date of the first day of her last period. Accurate to $\sim \pm 14$-21 days. Affected by variation in menstrual cycles (e.g. with under-nutrition or after cessation of hormonal contraceptive methods, or breastfeeding), socio-cultural attitudes to menstruation, literacy levels and digit preference.</p> <p><i>Best obstetric measure</i>: Combination of ultrasound and LMP.</p> <p><i>Fundal height</i>: Symphysis-fundal height measurement is another method that can estimate gestational age before birth. WHO guideline development group does not recommend this method due to lack of evidence in accuracy [7].</p>
Measuring gestational age after birth	There are a range of clinical assessments, e.g. Dubowitz, Ballard, Finnstrom scores and others [8, 9]. This study does not apply these methods.

economic status, limited literacy, high parity, and younger age are associated with increased uncertainty regarding LMP [24]. Other commonly used surrogates for GA measurement are described in Table 1 [21, 25, 26].

Measurement of child health and pregnancy outcomes in high burden countries which account for around two-thirds of the world's births still rely mainly on large-scale household surveys like Demographic and Health Surveys (DHS) rather than on civil and vital registration or routine health management information systems (HMIS) [11]. Most surveys, including DHS, do not include questions on GA for livebirths. However, questions which use women's report of GA are asked for non-livebirths in DHS to classify stillbirths, and for neonatal deaths in verbal autopsy tools [27, 28].

To our knowledge, no study has so far assessed GA questions to add to a survey such as DHS, and compared these against a gold standard early pregnancy 'ultrasound measurement'.

This paper is part of a series of papers from the Every Newborn-International Network for the Demographic Evaluation of Populations and their Health (EN-INDEPTH) study in five health and demographic surveillance system (HDSS) sites in Africa and Asia. This paper addresses three objectives:

1. Investigate completeness and feasibility of recording GA data in months and weeks by women's report in the EN-INDEPTH population-based survey in five HDSS sites using new/modified questions, including predictors of reporting.
2. Compare accuracy of GA reported in the EN-INDEPTH survey to GA recorded through

prospective health and demographic surveillance (Bandim and Matlab sites) and to GA assessed through early pregnancy ultrasound (Matlab site).

3. Undertake qualitative research to assess community perceptions, practices and barriers to reporting GA in population-based surveys, and identify commonalities and differences across the sites

Methods

EN-INDEPTH study design and settings

The EN-INDEPTH study was a cross-sectional multi-site study conducted between July 2017 and August 2018, including a survey of 69,176 women aged 15-49 years in five HDSS sites: Bandim in Guinea-Bissau, Dabat in Ethiopia, IgangaMayuge in Uganda, Matlab in Bangladesh and Kintampo in Ghana (Fig. 1). The protocol and main study paper are published elsewhere and provide further details [29, 30]. The primary objective of the study was to compare two methods of retrospective recording of pregnancy outcomes in surveys: full birth history with additional questions on pregnancy losses (FBH+), and full pregnancy history (FPH) as detailed elsewhere [29, 30].

Both woman and interviewer data were collected on Android tablets using the Survey Solutions data collection and management system [31]. Interviewers were recruited locally and were familiar with the culture and dialect of the study area. Following completion of data collection, data from the five HDSS sites were anonymised by local HDSS scientists, encrypted and then shared [29]. Data management and analysis were done using Stata version 15.1. Results are reported in accordance with STROBE Statement checklists for cross-sectional studies [32] (Additional file 1).

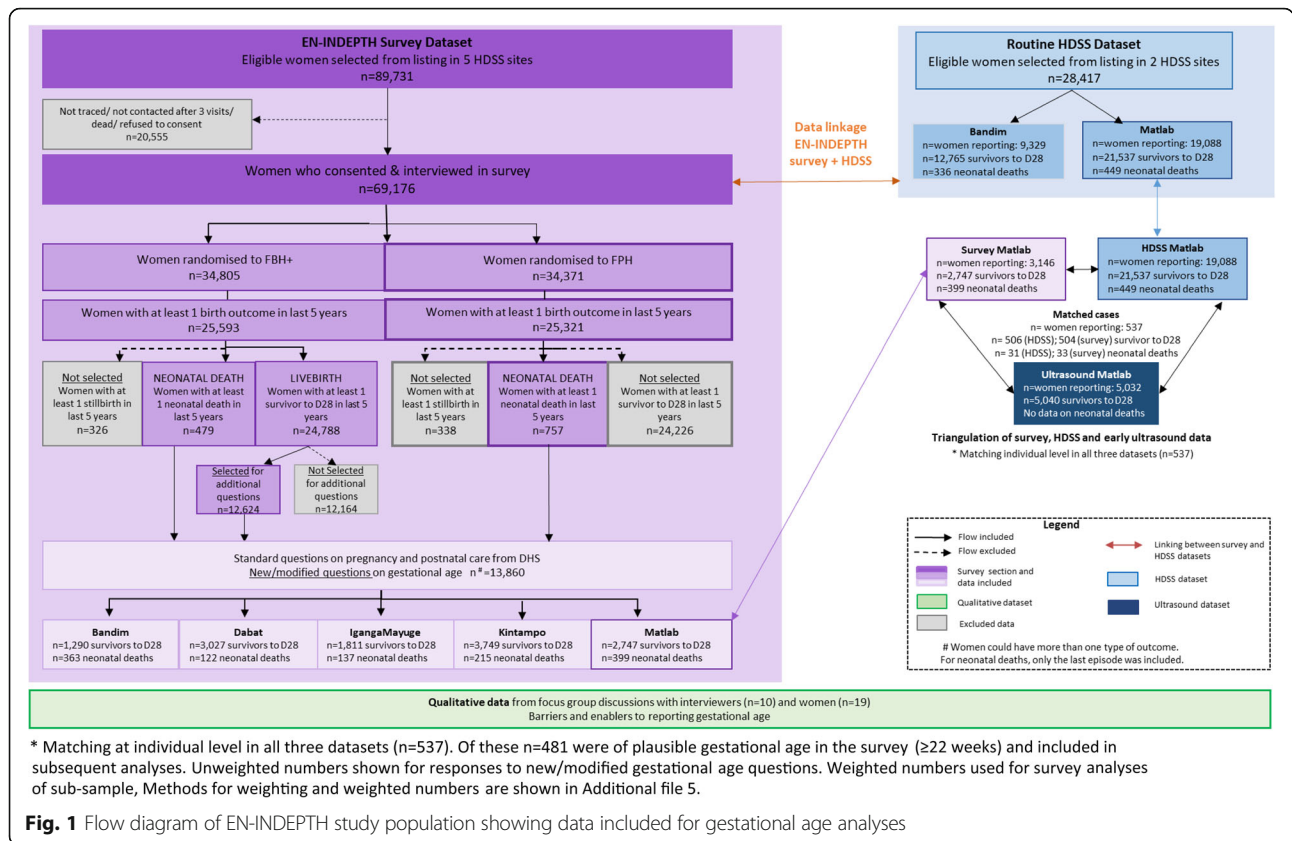


Fig. 1 Flow diagram of EN-INDEPTH study population showing data included for gestational age analyses

Focus group discussions (FGDs) with survey respondents and interviewers, and a survey of interviewers were performed in March-August 2018 [33]. Information on perceptions, practices, and barriers relating to knowledge and reporting of GA was collected. Qualitative data were transcribed using a combination of notes and audio recordings, and were coded and analysed using the qualitative data analysis software, NVivo 12.

Survey questions and HDSS linkage for gestational age

The EN-INDEPTH study also investigated the performance of existing or modified survey questions to capture other pregnancy-related outcomes including GA (Table 2). GA reported in months (GAM) was collected for all livebirths in the 5 years preceding the EN-INDEPTH survey. A sub-sample of survey respondents in all sites were also asked to report GA in weeks (GAW), and if they were ‘born before expected’ the number of weeks early for their most recent surviving livebirth, and all neonatal deaths in the last 5 years (Additional file 2). The two-part question on the woman’s perspective of whether her baby was ‘born before expected’, was adapted from the 2007 version of WHO’s Verbal Autopsy tool [28]. GAW

was collected from health cards where early was collected by recall only. For babies reported to be ‘born before expected’, GA in weeks was estimated as 40 minus the number of weeks early. A livebirth of GA < 9 months or GA < 37 weeks was coded as a preterm birth. Livebirths with reported GA ≤ 5 months or GA ≤ 21 weeks were excluded as survival below these limits is biologically implausible.

The EN-INDEPTH survey data were linked with HDSS data in the two sites where dates of LMP (Matlab, Bangladesh), and reported months of pregnancy at pregnancy registration (Bandim, Guinea-Bissau) were routinely recorded along with pregnancy outcomes (Additional file 3). In Matlab, ultrasound data from icddr, Matlab Hospital (Additional file 3) were also linked [29]. For these two sites, individual pregnancy records included in the EN-INDEPTH study since the 1st January 2012 were matched with that in the HDSS records using probabilistic matching (Additional file 4). Matlab Hospital records HDSS IDs in clinical records, enabling the matching of the ultrasound report with HDSS records. After probabilistic linking of births captured in survey with births in the HDSS, the matched children’s HDSS IDs were used to match ultrasound

Table 2 EN-INDEPTH survey questions for gestational age and method of administering these questions

Standard questions in DHS-7		New/modified questions asked on GA in EN-INDEPTH		
Question	Details	Question	Details	Probing, calculation at data collector's end, and recording
'How many months pregnant were you when that pregnancy ended?'	Only asked for non-livebirths. Asked in additional questions on non-livebirths after the roster (section 2).	'How many months did this pregnancy last?'	Asked for all births and pregnancy losses in FBH+ and FPH in roster. Acceptable range 0-11. Red error message shown if value outside range entered.	- Probed when women could not spontaneously report. - GA was recorded in completed months, e.g. a reported GA of 9 months 2 weeks was recorded as 9 months.
		'How many weeks pregnant were you when xxx was born?'	Asked for all last neonatal deaths and stillbirths since the 1st Jan 2012, and subset of surviving livebirths in section 4. Acceptable range 0-45. Red error message shown if value outside range entered. Data were collected from health cards (ANC card, maternal card or others) where it was available to women, else from recall.	- Most women reported GA in months. Then they were asked whether it was exactly the reported complete months, or the pregnancy lasted few more days/weeks, or that ended few days/weeks before the reported month. - Women's reported GAs in months were converted into weeks and recorded in completed weeks at interview. - Months and fraction of months (in days/weeks) reported after complete months were converted to weeks considering 30 days month in Matlab. - Bandim site did not probe. When women replied 'don't know' the response was recorded as such.
		'Was xxx born before expected?'	Asked for all last neonatal deaths and stillbirths since the 1st Jan 2012 and subset of surviving livebirths in section 4. Responses coded—Yes/No/Don't Know. Those responding 'Yes' were asked: 'How many weeks was xxx born before the expected date of delivery?' Response—numeric integer. Option to report in months if weeks are not known.	- Probed when women could not spontaneously report - Recorded in completed weeks - Recorded '0' for < 7 days - For months, recorded as the women reported

records. Only early ultrasound pregnancy dating reports at < 24 weeks were included in GA analyses [12].

Data analyses

Objective 1: completeness and feasibility of recording GA data in population-based surveys

For analyses of GAW and 'born before expected' questions, sample weights were applied using the *svyset* command to account for the different probability of a neonatal death being included compared to a livebirth surviving the neonatal period, given that women's response may vary for these two groups (Additional file 5). Descriptive statistics were used to analyse responses (any/plausible response) and digit/number preference for GA questions. Logistic regression was used to examine evidence of variations in GAW reporting (reporting any value against not reporting or reporting 'don't know') by socio-demographic characteristics and maternal care-seeking behaviour. Preterm birth rates were calculated for each approach and compared to

national estimates to assess plausibility of GA responses at a population level.

Century month code, DHS's date data coding system that uses month and year, was used to identify events occurring in the 5 years prior to the interview. Socioeconomic wealth quintiles were used to measure the wealth status of households and were derived from infrastructure, housing and assets owned using Principle Components Analysis as used by DHS and MICS [34].

Objective 2: accuracy of survey reported GA compared to routine HDSS and ultrasound data

GAW was calculated from HDSS data (Bandim and Matlab) and ultrasound data (Matlab only) (Additional file 3). In view of missing GAW in survey data from Bandim, survey GAM was compared with GAW from HDSS. In Matlab, GAW from the survey was compared to HDSS and ultrasound data (gold standard), and GAW from HDSS with ultrasound data. We categorized GAW in four groups (extreme and very preterm, $22 \leq \text{GAW} \leq 31$; moderate preterm, $32 \leq \text{GAW} \leq 36$; term and

post-term, $37 \leq \text{GAw}$) and then compared the groups based on GA estimates from HDSS and survey with the groups based on GAw from ultrasound. Sensitivity and specificity of preterm birth detection by GAw from the survey and HDSS were assessed. Bland-Altman mean difference (MD) between sources with 95% limits of agreement, concordance correlation coefficients (CCC) with 95% confidence interval (CI), and kappa coefficients (KC) with 95% CI were used to assess agreement. We used multinomial logistic regression to examine over- and under-reporting of GAw in survey and HDSS compared to ultrasound.

Objective 3: qualitative research to assess barriers and enablers to survey reported GA

To understand community perceptions and barriers related to GA reporting in household surveys, 29 focus group discussions (FGDs) were undertaken with 172 survey respondents and 82 survey interviewers and supervisors (Additional file 6) [33]. Thematic analysis to identify community perceptions, practices and barriers to reporting GA was conducted in NVivo 12 using an iterative process guided by an a priori codebook and addition of new codes that emerged during analysis. Themes were summarised and grouped to explore how findings contribute to understanding of the measurement of GA in population-based surveys.

Results

Overall

Information on GAM was collected for 65,562 livebirths in the last 5 years from 69,176 surveyed women. For the subsample of 13,860 livebirths, GAw and mother's perception of whether the child was born before the expected date was also collected (weighted number 15,086) (Fig. 1). Survey respondents differed across HDSS with regards to age, parity, education and religion (Table 3).

Objective 1: completeness and feasibility of recording GA data in population-based surveys

Completeness and plausibility of GA data captured in months

Table 4 panel A shows near-universal reporting of GAM for livebirths in the last 5 years in all five sites. However, in all sites except Matlab, 91-99% of babies were reported to have been born at 9 months (Fig. 2).

Completeness and plausibility of GA data captured in weeks

Completeness of GAw data was highly variable across sites (Table 4: panel B). In IgangaMayuge, 98.0% of women reported GAw compared to just 5.5% in Bandim. There were also reporting variations in GA in weeks by background characteristics (Additional file 7.1A). Reporting of GAw was higher amongst women above

30 years in Bandim; lower in women with ≥ 3 parity in Bandim and >3 parity in IgangaMayuge; higher in women who had ever attended school in Matlab; lower in highest wealth quintile in Bandim, in highest two wealth quintiles in Dabat and second to fourth wealth quintiles in Kintampo; lower in women affiliated with religions other than Islam and Christianity; lower amongst women who received 4+ ANC in Dabat and higher in 4+ ANC-receiving women in IgangaMayuge. Variations were not found by place of delivery.

Amongst those who reported to GAw, nearly all women reported a plausible value ($\text{GAw} \geq 22$) in all sites except Matlab (Table 4: panel B). In Matlab, 9.8% of the births reported GAw ≤ 21 weeks including identical GAM and GAw for 8%. Half of the 8% was reported by one interviewer who recorded the same values for GAw and GAM in 160 out of 171 records. Another 1.6% was reported by two interviewers and 2.4% by the other 17 interviewers. Subsequent analyses excluded births with $\text{GAw} \leq 21$.

Few livebirths were reported at < 36 weeks in any site (Fig. 3). In all sites, a preference for even digits was observed, with GAw heaped at 36 weeks (equalling to 4×9 months, the most commonly recorded value for GAM) in four sites (Additional file 7.1B). The questions on GAw were designed to collect GAw from card (ANC/other health cards) where it was available, else from women's recall. Of the GAw collected in the survey, 52% in Kintampo, 13% in IgangaMayuge and 0% in the other sites were from cards. Of the GAw from cards, $< 2\%$ were ≤ 21 weeks. Greater variation in reported GAw was seen by card compared to recall. A higher proportion of births were reported at 38 weeks in both sites, and fewer births reported at 36 weeks in Kintampo by card compared to recall (Fig. 4).

Other questions regarding preterm birth

Over 96.3% of women answered the question 'was xxx born before expected?' (Table 4: panel C). The proportion of 'Don't know/missing' responses to whether the baby was 'born before expected' or 'Don't know' to how many weeks the baby was 'born before expected' was 9.8% in Bandim and below 4% in other sites. The proportion of babies reported to have been 'born before expected' was 28.3% in Matlab, but only 1.4-5.4% in other sites.

Estimated preterm birth rates based on the three survey approaches tested

The estimated preterm birth rate using GAM was 17.0% in Matlab, compared with $\leq 3\%$ in all other sites (Table 4: panel A). GAw showed a similar preterm birth rate in Matlab (20.9%) but high rates in other sites (Dabat, 96.6%; Bandim, IgangaMayuge and

Table 3 Characteristics of women with at least one birth outcome in the last 5 years, EN-INDEPTH survey ($n = 50,914$)

	Bandim	Dabat	IgangaMayuge	Kintampo	Matlab	All sites
Number of women interviewed	9109	6254	5632	11,411	18,508	50,914
Age						
15-19	4.8	3.8	6.0	2.6	5.2	4.5
20-24	22.6	18.2	21.7	15.1	27.9	22.2
25-29	26.6	25.7	24.9	21.7	29.8	26.3
30-34	23.2	21.0	20.7	24.1	23.6	23.0
35-39	15.2	17.8	14.7	20.4	10.7	15.0
40-44	6.1	9.7	8.9	11.6	2.5	6.8
45-49	1.7	3.7	3.2	4.5	0.3	2.2
Total	100%	100%	100%	100%	100%	100%
Parity						
1	27.1	17.3	19.0	16.6	32.2	24.5
2	23.1	16.9	16.0	18.0	35.8	25.0
3+	49.8	65.8	65.0	65.4	32.0	50.5
Total	100%	100%	100%	100%	100%	100%
Education						
Never attended school	29.1	61.6	8.4	39.3	4.4	24.1
Primary incomplete	43.5	17.8	48.7	22.8	11.3	24.6
Primary complete	13.6	3.7	9.7	5.1	8.7	8.2
Secondary incomplete	13.7	16.0	33.3	32.9	64.5	38.9
Secondary complete/above	0.1	0.9	0.0	0.0	11.2	4.2
Missing	0.0	0.0	0.0	0.0	0.0	0.0
Total	100%	100%	100%	100%	100%	100%
Wealth index						
Lowest	20.2	26.5	25.3	20.3	20.4	21.6
Second	20.1	19.3	21.6	20.0	19.9	20.1
Middle	20.0	20.8	18.7	20.0	19.8	19.9
Fourth	19.4	17.8	18.7	19.9	20.1	19.5
Highest	20.3	15.6	15.8	19.9	19.9	19.0
Total	100%	100%	100%	100%	100%	100%
Religion						
Muslim	39.5	3.3	57.0	31.8	88.4	53.0
Christian	40.5	96.7	42.9	62.3	0.0	37.8
Other or none	20.0	0.0	0.2	5.9	11.6	9.1
Missing	0.1	0.0	0.0	0.0	0.0	0.0
Total	100%	100%	100%	100%	100%	100%

Kintampo, 59.5-71.5%) (Table 4: panel B). The question 'was xxx born before expected?' provided lower preterm birth rates GAM and GAW in Matlab (7.7%), and similar rates to GAM in the other sites (0.8-2.8%) (Table 4: panel C). Preterm birth estimates from all three survey approaches tested were very different from national estimates in all sites apart from Matlab (Additional file 7.4).

Objective 2: accuracy of survey reported GA compared to routine HDSS and ultrasound data

As only 5.7% of livebirths reported survey-GAW, estimated HDSS-GAW were compared to survey-GAM. HDSS-GAW was available for 5725 livebirths out of 13,456 livebirths with GAM ≥ 6 in the survey. Estimated GAW in the HDSS is almost normally distributed with an estimated preterm birth rate of 30.9%. In total, 93.2% of reported GAM were

Table 4 GA in months, weeks and 'born before expected' in the last 5 years, EN-INDEPTH survey

	HDSS sites					All sites
	Bandim	Dabat	IgangaMayuge	Kintampo	Matlab	
Number of women	9109	6254	5632	11,411	18,508	50,914
Panel A: Gestational age in months for livebirths in the last 5 years preceding start of survey						
Number of livebirths in the last 5 years	12,002	8295	8508	15,758	20,999	65,562
Implausible (≤ 5 months)	0.2	0.1	0.4	0.3	0.4	0.3
Plausible (≥ 6 months)	99.4	99.9	99.6	99.5	99.5	99.5
Do not know/missing	0.4	0.0	0.0	0.2	0.2	0.2
Total	100%	100%	100%	100%	100%	100%
Preterm birth rate ^a (CI)	1.7 (1.5, 1.9)	0.8 (0.6, 1.0)	3.0 (2.7, 3.4)	1.4 (1.2, 1.6)	17.0 (16.5, 17.5)	8.0 (6.4, 6.8)
Panel B: Gestational age in weeks for last livebirths in the last 5 years preceding start of survey						
Number of last livebirth in sub-sample (weighted)	2817	1978	2011	3629	4650	15,086
Implausible (≤ 21 weeks)	0.4	0.8	1.4	0.8	9.8	3.6
Plausible (≥ 22 weeks)	5.1	91.4	96.6	50.7	89.8	65.7
Do not know/missing	94.5	7.8	2.0	48.5	0.3	30.7
Total	100%	100%	100%	100%	100%	100%
Preterm birth rate ^a (CI)	59.5 (50.5, 67.1)	96.6 (95.6, 97.4)	69.4 (67.3, 71.4)	71.5 (69.4, 73.6)	20.9 (19.7, 22.2)	54.2 (53.2, 55.2)
Panel C: Babies 'born before expected' for livebirths in the last 5 years preceding start of survey^b						
Number of last livebirth in sub-sample (weighted)	2817	1978	2011	3629	4650	15,086
Implausible (≤ 21 weeks)	0.3	0.1	1.2	0.2	0.1	0.3
Plausible (≥ 22 weeks)	88.3	98.6	98.4	95.6	98.8	96.0
Do not know/missing whether the baby was born before expected	9.8	1.2	0.3	3.8	0.4	3.1
The baby was born before expected but do not know how many weeks before expected	1.6	0.0	0.2	0.4	0.4	0.6
Total	100%	100%	100%	100%	100%	100%
Preterm birth rate ^c (CI)	1.0 (0.7-1.5)	0.8 (0.5-1.3)	2.8 (2.1-3.6)	1.7 (1.3-2.2)	7.7 (6.9, 8.5)	3.5 (3.2-3.8)

^aPreterm birth rate was estimated amongst those who provided valid GA: GAm ≥ 6 and GAw ≥ 22

^bGAw was measured by subtracting number of weeks the child was 'born before expected' from 40 weeks. The survey did not ask when the baby was expected to be born. We have assumed 40 weeks for all babies

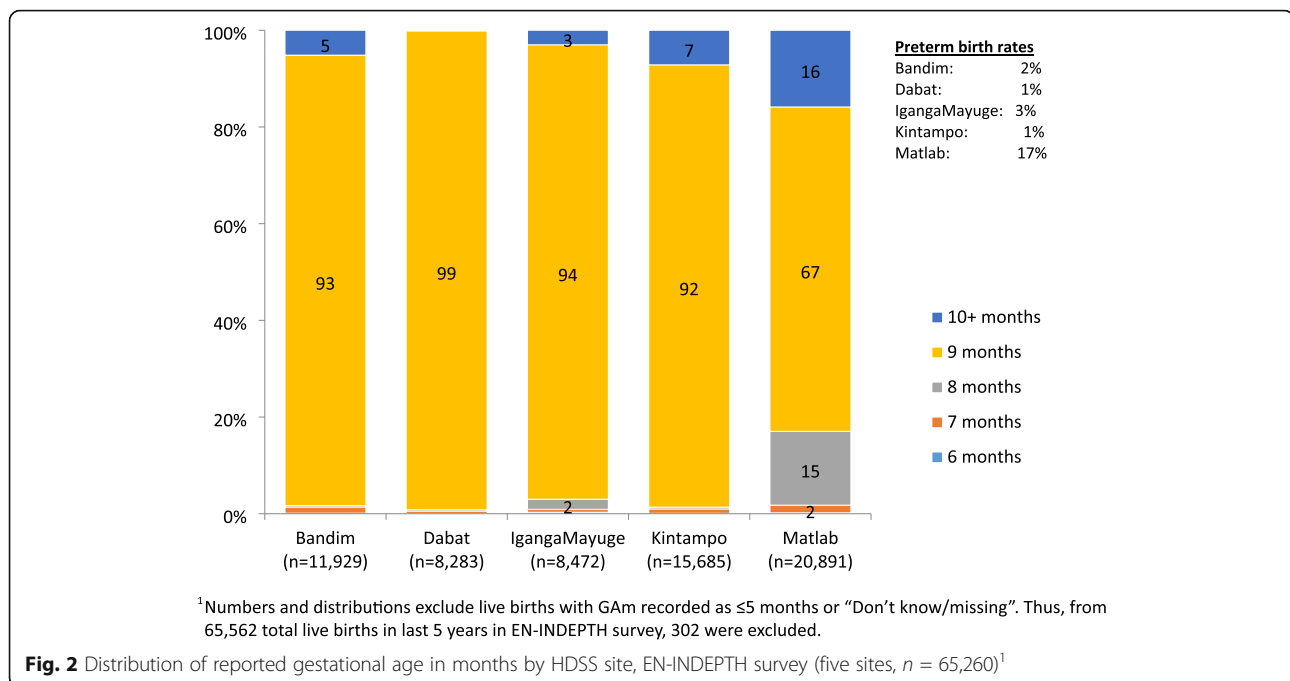
^cPreterm birth rate: GAw ≤ 36 , like panel B. All 'Don't know/missing' are excluded from the estimation

heaped at 9 months with 5.3% at 10 months, and a very low estimated preterm birth rate of 1.3% (Fig. 5a).

In Matlab, data from 2776 of 2907 with GA ≥ 22 weeks in the EN-INDEPTH survey were matched with HDSS data. Figure 5b shows the GAw distribution where HDSS-GAw peaked at 39-40 weeks and survey-GAw at 38 weeks. The estimated preterm birth rate was 12.9% by HDSS-GAw and 22.1% by survey-GAw. A total of 1079 of 2907 livebirths in the survey were matched to ultrasound estimated GA, 542 of these were excluded as occurred at ≥ 24 weeks. The 537 ultrasound reports before 24 weeks were matched to HDSS and survey data. The quality of GA data for these 537 cases is shown in Fig. 6. Subsequent analyses include only the 481 livebirths with GA ≥ 22 weeks in the survey (Fig. 5c). HDSS-GAw had a similar number of livebirths reported at 38, 39 and 40 weeks. Ultrasound GAw peaked at 39 weeks. HDSS-GAw estimated more after 39 weeks and less before 37

weeks than ultrasound GAw. This resulted in a slightly lower estimated preterm birth rate in the HDSS (12%) than ultrasound (14%) (Fig. 5c). The survey GAw tended to heap on even numbers. Heaping on 36 weeks may explain the higher estimated preterm birth rates with survey GAw compared to HDSS GAw and ultrasound GAw (see Additional files 7.2A-7.2C).

Agreement by simple group-to-group matching of categorical data (extreme/very preterm, moderately preterm, term and post-term) between HDSS GAw and ultrasound GAw was 87.3%, and 71.5% between survey GAw and ultrasound GAw (Table 5A). However, the overall agreement between ultrasound GAw and HDSS GAw was weak (kappa coefficient (KC) = 0.54), and was poor between ultrasound GAw and survey GAw (KC = 0.25) [35]. For a simpler grouping (term and preterm), the agreement improved to 0.65 (KC) between HDSS-GAw and ultrasound-GAw, and to 0.36 (KC) between



survey-GAw and ultrasound-GAw. Bland-Altman mean difference (MD) and concordance correlation coefficients showed similar results with better agreement between ultrasound GAw and HDSS GAw than ultrasound GAw and survey GAw (Fig. 7).

Validity of GA data in HDSS and survey compared to gold standard ultrasound GA data in Matlab

Of the GAw linked amongst ultrasound, HDSS and survey, 38.3% in HDSS and 20.4% in the survey had an exact match to ultrasound. Over reporting of GAw in both HDSS and survey was around one in three. Close to half (44.7%) of GAw in the survey were under reported.

Results from multinomial logistic regression did not find any variations in over- or under- reporting of GAw in the survey compared to ultrasound GAw by background characteristics. Lower over- reporting of HDSS GAw compared to ultrasound GAw was seen in the middle to fourth wealth quintiles, and higher over-reporting was observed in primary educated women. Higher under reporting was found in non-Muslims and primary educated women. Women’s age, parity, TV watching, ANC visits, place of delivery, icddr,b service area and survey recall period were not associated with over- or under- reporting (Additional file 7.2D).

The sensitivity of using HDSS collected GAw to detect preterm birth was 66%, and specificity was 95% compared to ultrasound ‘gold standard’ (Table 5B). Similar patterns with slightly lower levels were seen for survey collected GAw, with 60% sensitivity and 93% specificity.

Objective 3: qualitative research to assess barriers and enablers to survey reported GA

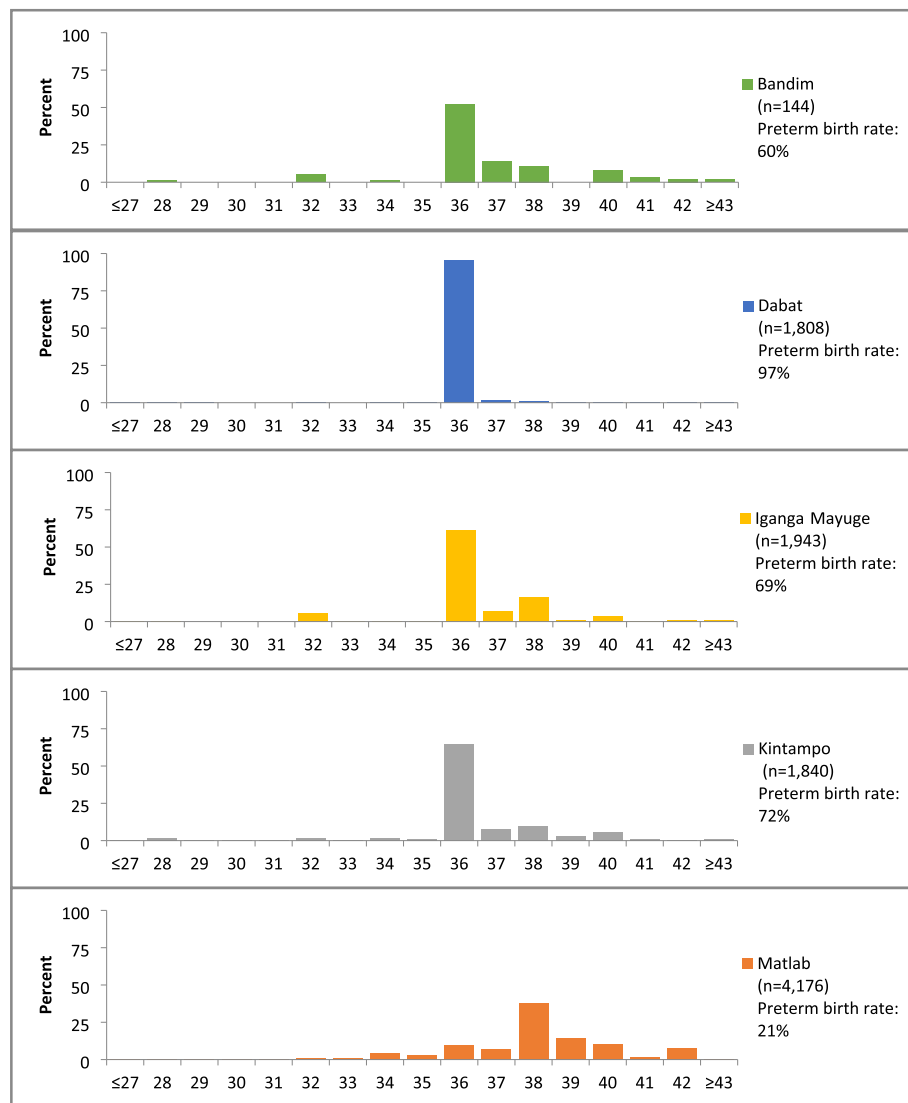
Community perceptions

Women perceived the importance of tracking the progress of pregnancy in all sites as this was seen to help in birth planning and preparation (Additional file 7.3). Facilitating fathers to be available to accompany the mother for ANC and delivery was another reason in IgangaMayuge. In Kintampo, women were scolded by healthcare providers if they could not report GA at ANC visits. Knowing the date of conception was also important in IgangaMayuge, especially for younger women to avoid denial of conception by the child’s biological father.

Community practices

Measuring or counting GA differed across sites. Women in Bandim found this difficult, whilst in IgangaMayuge it was perceived as easy. Women in IgangaMayuge and Kintampo reported that the ANC provider helped calculate GA.

Women tended to count GA in months in all sites. Missed periods, religious and cultural events, crop harvesting months and other key time points were used as reference points to count the months. Women in Dabat used key events in their religious calendar to recall their LMP date. For example, one woman stated, ‘my menstruation was terminated at Yetir Mariam’, meaning 21 January. In Matlab and Kintampo, women reported counting GA by missed periods—some counted their first missed period as the first month of GA, whilst



¹ Numbers and distributions exclude live births with GAw recorded as ≤ 21 weeks or "Don't know/missing". Excluded n=5,174.

Fig. 3 Distribution of reported gestational age in weeks by HDSS site, EN-INDEPTH survey (fives sites weighted, $n = 9912$)¹

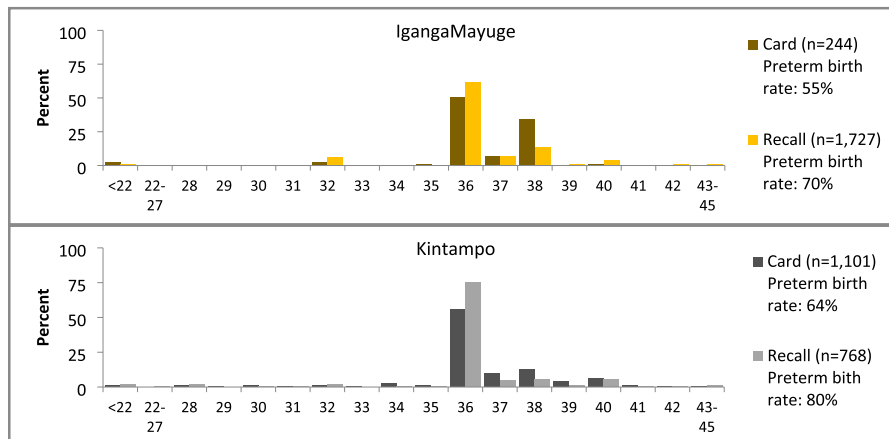
a few others counted the first month of GA as their second missed period. GA counting in Matlab varied by religious affiliation. Hindu women counted 10 months 10 days for a full-term pregnancy whilst Muslim women counted 9 months.

The Hindus usually tell like ten months ten days. In contrast the Muslims tell, 'it remains nine months, does it exceed nine months?' I worked mostly with the Hindus. I got ten months ten days from them, though I probed them well. Despite probing, they said ten months ten days. (Interviewer, Matlab, Bangladesh)

Reported barriers and enablers

Women's education (in Kintampo, Matlab and Dabat), and ANC attendance or facility birth (in Bandim) were perceived to improve GA reporting.

Barriers to knowing LMP included conceiving before their menses had returned following a previous pregnancy, cessation of hormonal contraceptives (Dabat and Matlab) or lack of awareness of menstrual cycles. Whilst health cards were perceived as a potential enabler, they were frequently poorly completed by healthcare workers and not preserved by many women. Social stigma and fear of witchcraft was an additional barrier to GA reporting in Bandim.



¹ GAw available for 1,971 out of 2,011 live births in last five years in IgangaMayuge, and 1,869 out of 3,629 in Kintampo (weighted numbers). Note: Preterm birth rates were estimated excluding GA<22 weeks

Fig. 4 Comparison of GA weeks from card and recall, EN-INDEPTH survey, IgangaMayuge (n = 1971) and Kintampo (n = 1869)¹

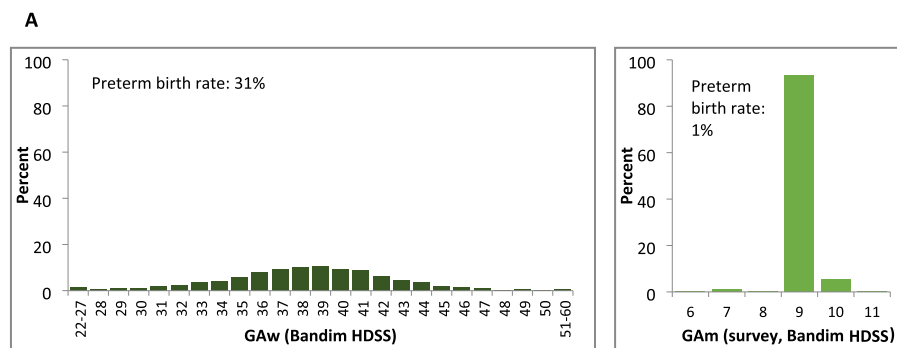
Some don't count their gestational days because of witchcraft; say, if SOMEONE else knows you are pregnant, he/she will be waiting for you at the birth on delivery day...sometimes someone is three or four months pregnant and still deny it and doesn't say anything. (Interviewer, Bandim, Guinea-Bissau)

Some interviewers reported specific issues in obtaining GA information in Matlab and Kintampo sites where probes were required to help women recall LMP and GA_m, and the interviewer then calculated GA_w themselves based on information provided by the respondents.

Discussion

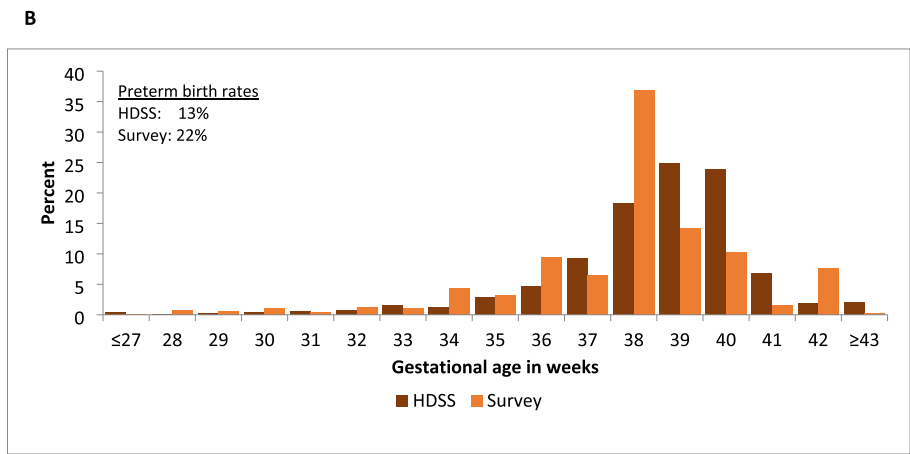
Given the high burden of deaths and disability-adjusted life years due to preterm birth, improving data on gestational age is a high priority, especially from the highest-burden countries where household

survey data remains a primary data source. To our knowledge, this is the first study to assess household survey questions on GA regarding feasibility, and importantly, validity compared with ultrasound-based GA as a gold standard in a subsample of the EN-INDEPTH study. Our findings in this large dataset from five countries suggest, whilst women can almost universally report GA_m, these results are severely heaped on 9 months, with resultant underestimation of preterm birth rates. Reporting of GA_w was feasible and of moderate sensitivity to detect preterm birth. In the other four sites, reporting of GA_w was highly variable in terms of both completeness and quality of reported data. Further investment is needed to overcome the barriers to collecting data on GA_w, and our study identifies some specific advances to improve the survey questions and the processes, underlining that addressing heaping is crucial.



¹ Survey-GA_m ≤5 and HDSS-GA_w ≤21 are excluded. From HDSS, 4 out of 5,725 GA excluded as were ≤21 weeks

Fig. 5A EN-INDEPTH survey GA weeks versus HDSS data GA months in the last 5 years, Bandim (n = 5721)¹



¹GAw≤21 are excluded. Matched cases in Matlab HDSS (n=2,776)

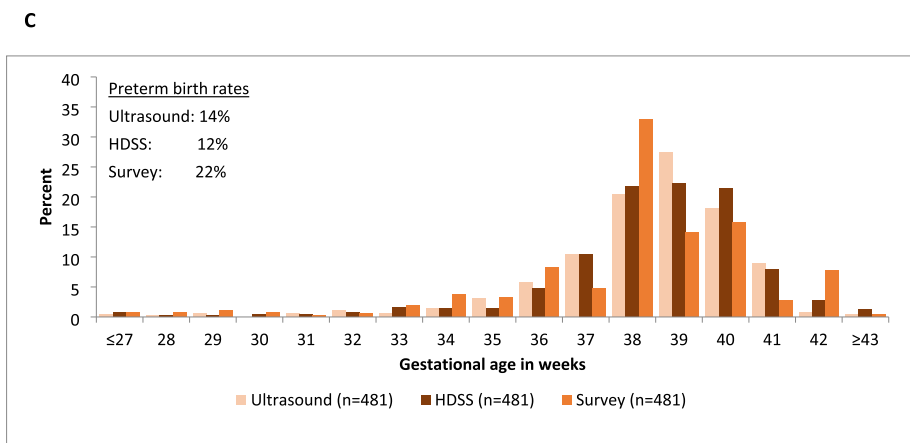
Fig. 5B EN-INDEPTH survey versus HDSS data for GA weeks in the 5 years, Matlab (n = 2776)¹

GAM was very feasible to answer, with almost 100% of women responding but in four of the five sites severe heaping on 9 months resulted in implausibly low estimated preterm rates (<3%) (Additional file 7.4) [1, 36]. Such heaping might be the result of women’s rounding up to the month of delivery or rounding by the interviewer. The exception to this was the Matlab site where GA_m produced an estimated preterm birth rate of 17.0%.

Reporting of GAW was highly variable and required probing to obtain a specific response. Probing was not used in Bandim and 94% of responses were recorded as ‘don’t know or missing’, whilst in Kintampo, 44% of responses were ‘don’t know or missing’, even after probing. In four sites, it seems from the GAW distribution that GAW was predominantly calculated by the interviewers multiplying GAM by four, resulting in

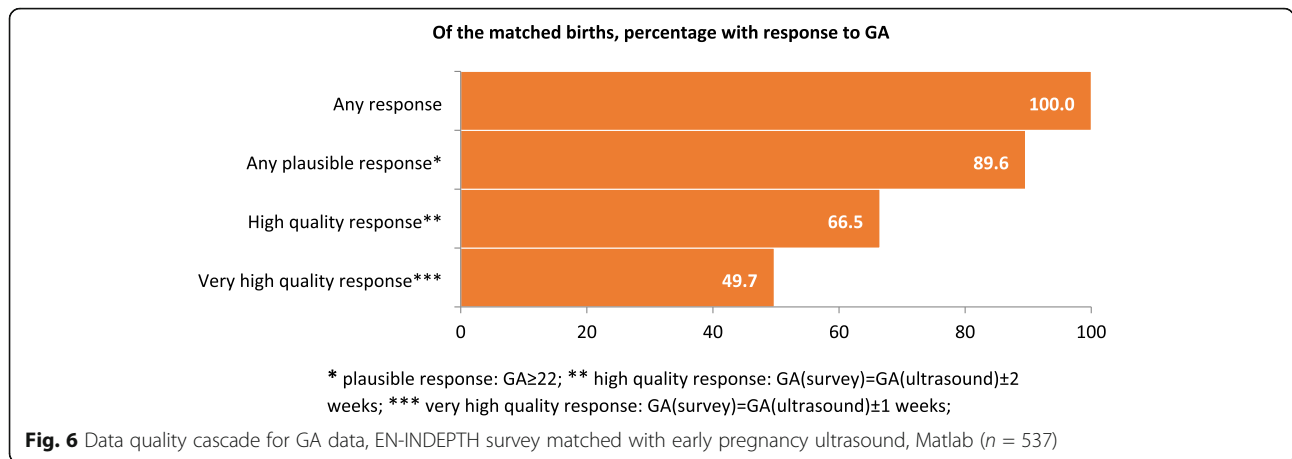
high estimated preterm birth rates (59.5-96.6%). In Matlab, data collectors were trained to multiply GAM by four and add 2 to get GAW, and to take into account any reported days or weeks before or after a completed month. This resulted in less heaping on 36 weeks, and an estimated preterm birth rate of 20.9%, which may still be an overestimate. Very few (<2%) of the reported GAW were implausible. Including in-built data quality checks for implausible responses, ≤ 21 weeks and implausible GAM/GAW combinations in future electronic data capture survey tools could reduce such errors. Further research is needed to test this approach in other settings.

A question to the woman, if her baby was ‘born before expected’, was adapted from the 2007 version of WHO’s Verbal Autopsy tool [28], and was feasible to answer but resulted in preterm birth rates which were implausibly low in all sites apart from in Matlab. Accurate answers



¹GAw≤21 are excluded. Matched cases (Matlab only), n=481. EN-INDEPTH survey estimate based on women’s report. HDSS data including early pregnancy testing.

Fig. 5C GA weeks (early pregnancy ultrasound) versus EN-INDEPTH survey, and HDSS data, last 5 years, Matlab (n = 481)¹



require the woman to know her expected date of delivery (EDD). Whilst EDD should be routinely calculated at first ANC visit, despite 2/3rd ANC coverage in Dabat and > 90% in all other sites, these data suggest that this information is not communicated to the woman, or she is unable to recall or unwilling to report it.

Analyses of GA amongst the births in EN-INDEPTH survey linked with Matlab hospital’s pregnancy ultrasound data show similar rates of preterm births in ultrasound and HDSS, but higher in the survey. Other studies have found LMP-based measurement tends to report higher GA than ultrasound-based measurement [17, 22]. We note that GAW patterns and socio-demographic characteristics were similar between the groups, but ANC seeking and facility delivery were

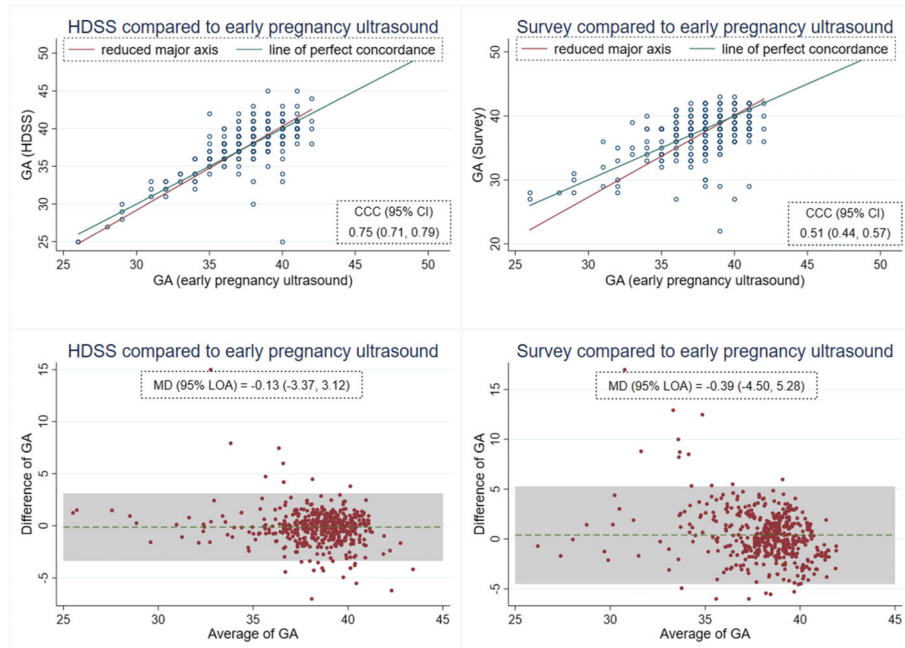
higher in matched group as the matched cases came from icddr,b service area. Over- or under- reporting of GAW in the survey compared to ultrasound were similar irrespective of women’s age, parity, TV watching, religion, dose-response of ANC care and place of delivery. All women amongst the matched group had at least one ANC in the last pregnancy, hence, status related to GA reporting remained unknown for women who did not receive any ANC. Survey GAW were over- reported amongst women with no education and under reported amongst women from second wealth quintile than ultrasound GAW.

Our qualitative data suggest that women track GA in pregnancy since this is perceived important to know to be able to plan for ANC and delivery, including getting

Table 5A Agreement for gestational age category between ultrasound, HDSS and survey (Matlab only, n = 481)

	Extreme/very preterm ^a	Moderate preterm ^b	Term ^c	Post-term ^d	n	Kappa [confidence interval]
Panel A: % of maturity group in HDSS matched with that of ultrasound (group-to-group match: 87.3%)						
HDSS	Ultrasound					0.54 [0.49, 0.54]
Extreme/very preterm	7	1	2	0		
Moderate preterm	2	34	13	0		
Term	0	22	378	3		
Post-term	0	1	17	1		
n	9	58	410	4	481	
Panel B: % of maturity-group in survey matched with that of ultrasound (group-to-group match: 71.5%)						
Survey	Ultrasound					0.25 [0.22, 0.30]
Extreme/very preterm	7	4	7	0		
Moderate preterm	2	27	56	1		
Term	0	27	309	2		
Post-term	0	0	38	1		
n	9	58	410	4	481	

^a22 ≤ GAw ≤ 31
^b32 ≤ GAw ≤ 36
^c37 ≤ GAw ≤ 41
^dGAw ≥ 42



¹Indicates the matched cases (Matlab only), n = 481. EN-INDEPTH survey estimate based on women's report. HDSS data include early pregnancy testing. CCC stands for concordance correlation coefficient, MD for Bland-Altman mean difference and LOA for limits of agreement

Fig. 7 GA weeks comparing early pregnancy ultrasound with EN-INDEPTH survey and with HDSS (n = 481)¹

the support of the father, and to be able to tell the health provider. Whilst almost all women in all sites were able to report GAM with plausible values, sometimes using religious and cultural events to assist recall, FGDs with women and interviewers suggested large variation in how women count 'months'. Variation

in reported length of gestation may be affected by cultural norms such as Matlab's Hindu women reporting GA as 10 months 10 days, biological differences such variation in length of menstrual cycles or conceiving after a period of amenorrhoea, or use of different calendars such as 30.4 days in a Gregorian calendar compared to 29.5 days in a lunar calendar. All these can impact on comparability of survey-captured GA.

Table 5B Validity of GA weeks data comparing ultrasound to HDSS and to EN-INDEPTH survey (Matlab, n = 481)

Panel A: Baby's maturity—HDSS compared to ultrasound (gold standard)					
		Ultrasound		n	Sensitivity (HDSS) = 65.7% Specificity (HDSS) = 94.5% Positive predictive value (HDSS) = 74.6% Negative predictive value (HDSS) = 96.4% Kappa [CI] = 0.65 [0.55, 0.76]
		Preterm ^a	Term ^b		
HDSS	Preterm	44	15	59	
	Term	23	399	422	
		n	67	414	481
Panel B: Baby's maturity—survey compared to ultrasound (gold standard)					
		Ultrasound		n	Sensitivity (survey) = 59.7% Specificity (survey) = 92.8% Positive predictive value (survey) = 38.5% Negative predictive value (survey) = 92.8% Kappa [CI] = 0.36 [0.26, 0.46]
		Preterm	Term		
Survey	Preterm	40	64	104	
	Term	27	350	377	
		n	67	414	481

^a22 ≤ GAw ≤ 36

^bGAw ≥ 37

Improving GA data from population-based surveys requires that women know the information, and this may be facilitated by paper-based tools such as calendars, or smart-phone apps, improved access to early ANC and ultrasound and communication from health workers. Women must also be able and willing to report this information at the time of the survey. Including this information in ANC or maternal-child health cards could facilitate data availability at the time of the survey. In some settings, such as Bandim, social stigma and fear of witchcraft may need to be addressed.

Handheld health cards, such as antenatal or child health records are potentially effective for communicating information from health providers to an interviewer in a household survey, and are commonly used to collect information on birthweight [37, 38]. Although cards were expected to be better sources for GAw, we found similar GA distributions between cards and recall in Kintampo and IgangaMayuge. This may be as many women first attend ANC in late pregnancy and health workers, hence,

rely on women's reported LMP or stated GA. Health cards were rarely used to report GA in other sites, which could be a missed opportunity, for example, in Bandim birthweight from card was available for 46.2% of livebirths compared to just 0.7% with GA from card [39]. GA-related information may vary by type of card and includes the expected date of delivery or GA at a visit or birth (in weeks or months). Processes used by interviewers to record GA from the information on the card are unknown, and the higher than expected estimated preterm birth rates could be explained by conversion from months.

This study has strengths and limitations. Strengths include the large survey dataset from five LMICs, with consistent questions and analyses, plus multi-site comparable, qualitative data. Linkage of the survey with HDSS and ultrasound data from Matlab is novel, however, the generalisability of these findings may be limited as women with early ultrasound may be systematically different from other women in Matlab, for example with higher care-seeking, and from women in other settings without intensive pregnancy surveillance with widespread early pregnancy testing.

Access to ultrasound use during pregnancy is increasing, for example, in 2017, 74% of recently delivered women in Bangladesh reported having had an ultrasound during pregnancy [40]. However, early ultrasound coverage is presumed to be lower. In this study, only about a half of 1079 matched women with a pregnancy ultrasound in Matlab had the ultrasound before 24 weeks. In addition to the challenge in accessing care early in pregnancy, costs and infrastructure requirements may impede widespread early pregnancy ultrasound scale-up in many settings [41]. Where early ultrasound is not feasible, LMP may be reasonably accurate, especially in societies where cultural restrictions placed on the undertaking of certain activities increase awareness of menstrual cycles [42]. Innovative solutions are required to facilitate women's full participation in society during menstruation, coupled with innovative methods to empower women to track their menstrual cycles. Prospective collection of LMP data alongside the use of a home calendar resulted in a high sensitivity (86%) and specificity (96%) for classifying preterm birth in Bangladesh [16].

Several of the challenges we identified regarding GA assessment in surveys are similar to those faced for birthweight in surveys, notably missing data, and heaping [38, 43]. Unlike GA, information on birthweight is routinely collected through household surveys and is sometimes used as a proxy for GA, although it is a poor proxy especially in South Asia where a high proportion of babies are born small for GA [3, 11]. In view of the importance of preterm and low birthweight outcomes, both GA and birthweight need further research to improve accuracy in survey data.

Based on these results, we propose a revised set of questions to collect GA information retrospectively in household surveys (Additional file 8). These questions focus firstly on collating prospectively collected data to inform GA from ultrasound or ANC card, and only asking women's retrospective report of length of gestation where no prospective data are available. These data could then be used by data collection apps during the survey or at the analysis stage standardise the calculation of GA.

Conclusions

Estimates of preterm birth rates based on GA can be feasible from population-based surveys. However, more work is needed to improve the accuracy of reported GA and would be best focused on improving the capture of information on pregnancy duration in weeks, using prospectively collected data from early pregnancy ultrasound or ANC visit records where available. We propose revised questions, and standardised probes which can be tested against gold standard early ultrasound data for validation.

Given the value of GA data and the major global data gaps for preterm birth estimates, further investments and innovations are justifiable to improve GA data in surveys. Importantly, whilst accuracy may be improved by better survey tools, a pre-requisite is that women know their menstruation dates. This will require a shift in social norms, both to reduce the stigma in discussing menses and improving women's awareness regarding the recording of dates.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12963-020-00230-3>.

Additional file 1. STROBE guidelines checklist.

Additional file 2. Selection of women with a livebirth surviving the neonatal period, EN-INDEPTH survey.

Additional file 3. Overview of gestational age data collection in Matlab and Bandim HDSS sites.

Additional file 4. Linking between EN-INDEPTH survey and HDSS data.

Additional file 5. Calculation of survey weights. 5.1: Methods for calculation of survey weights. 5.2: Weighted numbers of livebirths by HDSS sites.

Additional file 6. Qualitative methods for Focus Group Discussions in the EN-INDEPTH study.

Additional file 7. Results: Additional details. 7.1: Gestational age capture in weeks in the EN-INDEPTH survey. 7.1A: Logit estimates of adjusted ORs and 95% confidence interval of responding to GA in weeks. 7.1B: Matching of GA months with GA weeks, EN-INDEPTH survey. 7.1C: Gestational age distribution by religion, Matlab site, EN-INDEPTH survey in last five years. 7.2: Comparison of GA weeks between survey, HDSS and early pregnancy ultrasound, Matlab site. 7.2A: GA weeks in last five years by HDSS, early pregnancy ultrasound and five years prior to EN-INDEPTH survey. 7.2B: GA weeks for livebirths by HDSS, early pregnancy ultrasound and EN-INDEPTH survey in last five years. 7.2C: Early pregnancy ultrasound

versus EN-INDEPTH survey and HDSS data in last five years by ultrasound timing. 7.2D: Adjusted relative risk ratios for over- and under-reporting of GA weeks, survey and HDSS versus ultrasound. 7.3: Community perceptions, practices and barriers to reporting GA, EN-INDEPTH study (five sites). 7.4: Comparison of preterm birth rates in EN-INDEPTH study to external data sources.

Additional file 8. Proposed revised questions to capture gestational age.

Additional file 9. Ethical approval of local Institutional Review Boards.

Abbreviations

CCC: Concordance correlation coefficient; CI: Confidence interval; DHS: Demographic and Health Survey; EN-INDEPTH: Every Newborn-International Network for the Demographic Evaluation of Populations and their Health; FBH+: Full birth history (+ denotes additional questions on pregnancy losses); FGD(s): Focus group discussion(s); FPH: Full pregnancy history; GA: Gestational age; GAm: Gestational age in months; GAw: Gestational age in weeks; HDSS: Health and demographic surveillance system; ID: Identification number; KC: Kappa coefficient; LMICs: Low- and middle-income countries; LMP: Last menstrual period; LOA: Limits of agreement; MD: Mean difference; MICS: Multiple Indicator Cluster Survey; WHO: World Health Organization

Acknowledgements

This supplement is dedicated to the memory of Professor Peter Byass, who was the Senior External Editor of the supplement. Peter died suddenly in August 2020 and will be greatly missed by the EN-INDEPTH study team and entire global health community.

We thank the 118 interviewers and many HDSS staff participating in this study for their hard work and dedication. Many thanks to Samuelina Arthur, Claudia DaSilva, Olivia Nakisita and the relevant site staff for their administrative support.

We express appreciation to the EN-INDEPTH expert advisory group: Fred Arnold; Peter Byass; Trevor Croft; Kobus Herbst; Sunita Kishor; Florina Serbanescu; Turgay Unalan; Shane Khan; Attila Hancioglu.

We acknowledge the core funders for all sites/institutions.

The International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), the EN-INDEPTH study implementing organisation in Bangladesh, gratefully acknowledges the institutional support of the Government of People's Republic of Bangladesh, Canada, Sweden, and the UK.

Finally, and most importantly, we thank the women participating in the EN-INDEPTH study and their families, without whom this work would not have been possible.

Ethics and consent to participate

The EN-INDEPTH study was granted ethical approval by the Institutional Review Boards in all operating countries as well as from the Institutional Ethical Review Committee of the London School of Hygiene & Tropical Medicine (Additional file 9). Respondents of every successful interview gave written consent/assent after being informed of the objective, data use, procedure of the interview, risks and benefits of participating in the study, right to withdraw from interview anytime point of time and not responding to questions where she feels discomfort. The study ensured the respondent's privacy at data collection and confidentiality at data use.

The every newborn-INDEPTH study collaborative group

Senior External Supplement Editors: Peter Byass; Stephen M Tollman; Hagos Godefay

Technical Supplement Editors: Joy E. Lawn; Peter Waiswa; Hannah Blencowe

Managing Supplement Editors: Judith Yargawa; Joseph Akuze (data and statistics)

Other EN-INDEPTH Collaborative Group Members:

By team: PI followed by other members in alphabetical order

Bandim: Ane B Fisker (PI); Justiniano SD Martins; Amabelia Rodrigues; Sanne M Thysen

Dabat: Gashaw Andargie Bikis (PI); Solomon Mokonnen Abebe; Tadesse Awoke Ayele; Telake Azale Bisetegn; Tadesse Guadu Delele; Kassahun Alemu Gelaye; Bisrat Misganaw Geremew; Lemma Dersseh Gezie; Tesfahun Melese;

Mezgebu Yitayal Mengistu; Adane Kebede Tesega; Temesgen Azemeraw Yitayew

IgangaMayuge: Simon Kasasa (PI); Edward Galiwango; Collins Gyezaho; Judith Kaija; Dan Kajungu; Tryphena Nareeba; Davis Natukwatsa; Valerie Tusubira

Kintampo: Yeetey AK Enuameh (PI); Kwaku P Asante; Francis Dzabeng; Seeba Amenga Etego; Alexander A Manu; Grace Manu; Obed Ernest Nettey; Sam K Newton; Seth Owusu-Agyei; Charlotte Tawiah; Charles Zandoh

Matlab: Nurul Alam (PI); Nafisa Delwar; M Moinuddin Haider; Md. Ali Imam; Kaiser Mahmud

LSHTM/ Makerere School of Public Health: Angela Baschieri; Simon Cousens; Vladimir Sergeevich Gordeev; Victoria Ponce Hardy; Doris Kwesiga; Kazuyo Machiyama

About this supplement

This article has been published as part of *Population Health Metrics Volume 19 Supplement 1, 2021: Every Newborn-INDEPTH study: Improving the measurement of pregnancy outcomes in population-based surveys*. The full contents of the supplement are available online at <https://pophealthmetrics.biomedcentral.com/articles/supplements/volume-19-supplement-1>.

Authors' contributions

The EN-INDEPTH study was conceptualised by JEL. All site teams contributed to the design of the study protocol and undertook data collection. MMH together with NA, HB and JEL developed the detailed research questions and overall analysis plan for this paper. These were refined with inputs from the wider EN-INDEPTH study collaborative group at a multi-country workshop in Entebbe Uganda in February 2019. DK, JM, QN, OEAN and AKT implemented the study at the sites and TA, SMTH, MAK provided the ultrasound data. JA prepared the survey data files. Quantitative data analysis was undertaken by MMH and KM, and qualitative analysis by MAI and ND. HB, SC and NA provided statistical oversight. JEL guided the objectives and plan of the manuscript. The manuscript was drafted by MMH, JEL, HB and MAI. SC, NA, VPH, JY and ABF provided valuable comments. All authors reviewed and agreed on the final version of the manuscript.

Funding

The EN-INDEPTH study (including publication costs) was funded by the Children's Investment Fund Foundation (CIFF) by means of a grant to LSHTM (PI *Joy E. Lawn*), and a sub-award to the INDEPTH MNCH working group with technical leadership by Makerere School of Public Health (PI *Peter Waiswa*).

Availability of data and materials

Data sharing and transfer agreements were jointly developed and signed by all collaborating partners. The datasets generated during the current study are deposited online at <https://doi.org/10.17037/DATA.00001556> with data access subject to approval by collaborating parties.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Health Systems and Population Studies Division, icddr,b, Dhaka, Bangladesh.

²Maternal, Adolescent, Reproductive & Child Health (MARCH) Centre, London School of Hygiene & Tropical Medicine, London, UK.

³Nutrition and Clinical Services Division, icddr,b, Dhaka, Bangladesh.

⁴Centre of Excellence for Maternal Newborn and Child Health Research, Makerere University, Kampala, Uganda.

⁵Department of Health Policy, Planning and Management, Makerere University School of Public Health, Kampala, Uganda.

⁶Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London, UK.

⁷Bandim Health Project, Bissau, Guinea-Bissau.

⁸Research Centre for Vitamins and Vaccines, Statens Serum Institut, Copenhagen, Denmark.

⁹Open Patient Data Exploratory Network (OPEN), University of Southern Denmark, Odense, Denmark.

¹⁰IgangaMayuge Health and Demographic Surveillance System, Makerere University Centre for Health and Population Research, Makerere, Uganda.

¹¹Matlab Health Research Centre, icddr,b, Dhaka, Bangladesh.

¹²Kintampo Health Research Centre, Kintampo, Ghana.

¹³Dabat Research Centre Health and Demographic Surveillance System, Dabat, Ethiopia.

¹⁴Department of Health Systems and Policy, University of Gondar Institute of Public Health, Gondar, Ethiopia.

Published: 8 February 2021

References

- Chawanpaiboon S, Vogel J, Moller A, Lumbiganon P, Petzold M, Hogan D, et al. Global, regional and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health*. 2019; 7:e37–46.
- Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn J, et al. Global, regional, and national causes of child mortality in 2000–13, with projects to inform post-2015 priorities: an updated systematic analysis. *Lancet*. 2015;385.
- Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A, et al. Born too soon: the global epidemiology of 15 million preterm births. *Reprod Health*. 2013;10:S2.
- Dbstet A. WHO: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. *Acta Obstet Gynecol Scand*. 1977;56:247–53.
- World Health Organization: International Classification of Diseases 10th revision (ICD-10). 2010. http://www.who.int/classifications/icd/ICD10Volume2_en_2010.pdf?ua=1 [Accessed May 2020].
- Howson CP, Kinney MV, Lawn JE. Born too soon : the global action report on preterm birth. New York: March of Dimes, PMNCH, Save the Children, World Health Organization; 2012.
- World Health Organization: WHO recommendation on symphysis-fundal height measurement. 2018. <https://extranet.who.int/rhl/topics/preconception-pregnancy-childbirth-and-postpartum-care/antenatal-care/who-recommendation-symphysis-fundal-height-measurement>. Accessed May 2020.
- Lee AC, Mullany LC, Ladhani K, Uddin J, Mitra D, Ahmed P, et al. Validity of newborn clinical assessment to determine gestational age in Bangladesh. *Pediatrics*. 2016;138:e20153303.
- Lee AC, Panchal P, Folger L, Whelan H, Whelan R, Rosner B, et al. Diagnostic accuracy of neonatal assessment for gestational age determination: a systematic review. *Pediatrics*. 2017;140:e20171423.
- Rosenberg R, Ahmed N, Ahmed S, Saha S, Chowdhury A, Black R, et al. Determining gestational age in a low-resource setting: validity of last menstrual period. *J Health Popul Nutr*. 2009;27:332–8.
- Lawn J, Gravett M, Nunes T, Rubens C, Stanton C. The GAPPS review group: global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. *BMC Pregnancy Childbirth*. 2010;10:S1.
- Butt K, Lim K, Bly S, Cargill Y, Davies G, Denis N, et al. Determination of gestational age by ultrasound. *J Obstet Gynaecol Can*. 2014;36:171–81.
- Chervenak FA, Skupski DW, Romero R, Myers MK, Smith-Levitin M, Rosenwaks Z, et al. How accurate is fetal biometry in the assessment of fetal age? *Am J Obstet Gynecol*. 1998;178:678–87.
- American College of Obstetricians and Gynecologists (ACOG): Methods for estimating the due date. Committee Opinion #700 2017.
- Wanyonyi SZ, Mariara CM, Vinayak S, Stones W. Opportunities and challenges in realizing universal access to obstetric ultrasound in sub-Saharan Africa. *Ultrasound Int Open*. 2017;3:E52–e59.
- Gernand AD, Paul RR, Ullah B, Taher MA, Witter FR, Wu L, et al. A home calendar and recall method of last menstrual period for estimating gestational age in rural Bangladesh: a validation study. *J Health Popul Nutr*. 2016;35:34.
- Hoffman CS, Messer LC, Mendola P, Savitz DA, Herring AH, Hartmann KE. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. *Paediatr Perinat Epidemiol*. 2008;22: 587–96.
- 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.
- Neufeld LM, Haas JD, Grajeda R, Martorell R. Last menstrual period provides the best estimate of gestation length for women in rural Guatemala. *Paediatr Perinat Epidemiol*. 2006;20:290–8.
- Pereira AP, Dias MA, Bastos MH, da Gama SG, Leal Mdo C. Determining gestational age for public health care users in Brazil: comparison of methods and algorithm creation. *BMC Res Notes*. 2013;6:60.
- Rosenberg RE, Ahmed AS, 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;27:332–8.
- Savitz DA, Terry JW Jr, Dole N, Thorp JM Jr, Siega-Riz AM, Herring AH. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *Am J Obstet Gynecol*. 2002;187:1660–6.
- Weinstein JR, Thompson LM, Diaz Artiga A, Bryan JP, Arriaga WE, Omer SB, et al. Determining gestational age and preterm birth in rural Guatemala: a comparison of methods. *PLoS One*. 2018;13:e0193666.
- Hall MH, Carr-Hill RA, Fraser C, Campbell D, Samphier ML. The extent and antecedents of uncertain gestation. *Br J Obstet Gynaecol*. 1985;92:445–51.
- Chang KT, Mullany LC, Khatri SK, LeClerq SC, Munos MK, Katz J. Validation of maternal reports for low birthweight and preterm birth indicators in rural Nepal. *J Glob Health*. 2018;8.
- Lee A, Panchal P, Folger L, Whelan H, Whelan R, Rosner B, et al. Diagnostic accuracy of neonatal assessment for gestational age determination: a systematic review. *Pediatrics*. 2017;140:e20171423.
- Demographic and Health Surveys: DHS model questionnaire- Phase 7. 2015. <https://dhsprogram.com/publications/publication-dhsq7-dhs-questionnaires-and-manuals.cfm> [Accessed May 2020].
- World Health Organization: Verbal autopsy standards: ascertaining and attributing causes of death. 2007. <https://www.who.int/healthinfo/statistics/verbalautopsystandards/en/index3.html> [Accessed May 2020].
- Baschieri A, Gordeev VS, Akuzé J, Kwesiga D, Blencowe H, Cousens S, et al. “Every newborn-INDEPTH” (EN-INDEPTH) study protocol for a randomised comparison of household survey modules for measuring stillbirths and neonatal deaths in five health and demographic surveillance sites. *J Glob Health*. 2019;9:010901.
- Akuzé J, Blencowe H, Waiswa P, Baschieri A, Gordeev VS, Kwesiga D, et al. Randomised comparison of two household survey modules for measuring stillbirths and neonatal deaths in five countries: the every newborn-INDEPTH study. *Lancet Global Health*. 2019;8:E555–66.
- World Bank. Survey solutions CAPI/CAWI platform: release 5.26. Washington DC: The World Bank; 2018.
- von Elm E, Altman DG, Egger M, Pocock SJ, Götzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335:806.
- Kwesiga D, Tawiah C, Imam A, Kebede A, Nareeba T, Enameh YA, Manu G, Beedle A, Fisker A, Waiswa P, et al. Barriers and enablers to reporting pregnancy and adverse pregnancy outcomes in population-based surveys: EN-INDEPTH study. *BMC Population Health Metrics* 2021;19(Supplement 1). <https://doi.org/10.1186/s12963-020-00228-x>.
- O’donnell O, Van Doorslaer E, Wagstaff A, Lindelow M. Analyzing health equity using household survey data: a guide to techniques and their implementation. Washington DC: The World Bank; 2008.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22:276–82.
- Villar J, Ismail LC, Victora CG, Ohuma EO, Bertino E, Altman DG, et al. 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;384:857–68.
- World Health Organization: WHO recommendations on home-based records for maternal, newborn and child health. 2018. <https://apps.who.int/iris/bitstream/handle/10665/274277/9789241550352-eng.pdf?ua=1> [Accessed May 2020].
- Blencowe H, Krusevec J, de Onis M, Black RE, An X, Stevens GA, et al. National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health*. 2019;7: e849–60.
- Biks GA, Blencowe H, Ponce Hardy V, Misganaw B, Angaw DA, Wagnew A, Abebe SM, Guadu T, Martins JSD, Fisker AB, Imam MA, Nettey OEA, Kasasa S, Di Stefano L, Akuzé J, Kwesiga D, Lawn JE. Birthweight data completeness and quality in population-based surveys: EN-INDEPTH study. *BMC Population Health Metrics*, 2021;19(Supplement 1).
- National Institute of Population Research and Training (NIPORT) and ICF International: Bangladesh Demographic and Health Survey 2017–18: key indicators. Dhaka and Rockville: National Institute of Population Research and Training (NIPORT) and ICF International; 2019.

41. Swanson D, Lokangaka A, Bauserman M, Swanson J, Nathan RO, Tshefu A, et al. Challenges of implementing antenatal ultrasound screening in a rural study site: a case study from the Democratic Republic of the Congo. *Glob Health Sci Pract.* 2017;5:315–24.
42. Wall LL, Teklay K, Desta A, Belay S. Tending the 'monthly flower': a qualitative study of menstrual beliefs in Tigray, Ethiopia. *BMC Women's Health.* 2018;18:183.
43. Channon AA, Padmadas SS, McDonald JW. Measuring birth weight in developing countries: does the method of reporting in retrospective surveys matter? *Matern Child Health J.* 2011;15:12–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

