Development and external validation of Indian population-specific Garbhini-GA2 model for estimating gestational age in second and third trimesters

Veerendra P. Gadekar,[a,b,h,j] Nikhita Damaraju,[a,b,j] Ashley Xavier,[a,b] Shambo Basu Thakur,[a,b] Ramya Vijayram,[a,b] Bapu Koundinya Desiraju,[c] Sumit Misra,[c] GARBH-Ini Study Group,[c,k] Nitya Wadhwa,[c] Ashok Khurana,[d] Swati Rathore,[e] Anuja Abraham,[e] Raghunathan Rengaswamy,[b,f,h] Santosh Benjamin,[e,g] Anne George Cherian,[g] Shinjini Bhatnagar,[c,l] Ramachandran Thiruvengadam,[c,i,*] and Himanshu Sinha[a,b,h,**]

[a]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India
[b]Centre for Integrative Biology and Systems medicinE, Indian Institute of Technology Madras, Chennai, India
[c]Maternal and Child Health Program, Translational Health Science and Technology Institute, Faridabad, India
[d]The Ultrasound Lab, Defence Colony, New Delhi, India
[e]Department of Obstetrics and Gynaecology, Christian Medical College, Vellore, India
[f]Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, India
[g]Department of Community Health, Christian Medical College, Vellore, India
[h]Robert Bosch Centre for Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, India
[i]Department of Biochemistry, Pondicherry Institute of Medical Sciences, Puducherry, India

The Lancet Regional Health - Southeast Asia 2024;25: 100362

Published Online 25 February 2024
https://doi.org/10.1016/j.lansea.2024.100362

Summary

Background A large proportion of pregnant women in lower and middle-income countries (LMIC) seek their first antenatal care after 14 weeks of gestation. While the last menstrual period (LMP) is still the most prevalent method of determining gestational age (GA), ultrasound-based foetal biometry is considered more accurate in the second and third trimesters. In LMIC settings, the Hadlock formula, originally developed using data from a small Caucasian population, is widely used for estimating GA and foetal weight worldwide as the pre-programmed formula in ultrasound machines. This approach can lead to inaccuracies when estimating GA in a diverse population. Therefore, this study aimed to develop a population-specific model for estimating GA in the late trimesters that was as accurate as the GA estimation in the first trimester, using data from GARBH-Ini, a pregnancy cohort in a North Indian district hospital, and subsequently validate the model in an independent cohort in South India.

Methods Data obtained by longitudinal ultrasonography across all trimesters of pregnancy was used to develop and validate GA models for the second and third trimesters. The gold standard for GA estimation in the first trimester was determined using ultrasonography. The Garbhini-GA2, a polynomial regression model, was developed using the genetic algorithm-based method, showcasing the best performance among the models considered. This model incorporated three of the five routinely measured ultrasonographic parameters during the second and third trimesters. To assess its performance, the Garbhini-GA2 model was compared against the Hadlock and INTERGROWTH-21st models using both the TEST set (N = 1493) from the GARBH-Ini cohort and an independent VALIDATION dataset (N = 948) from the Christian Medical College (CMC), Vellore cohort. Evaluation metrics, including root-mean-squared error, bias, and preterm birth (PTB) rates, were utilised to comprehensively assess the model's accuracy and reliability.

Findings With first trimester GA dating as the baseline, Garbhini-GA2 reduced the GA estimation median error by more than three times compared to the Hadlock formula. Further, the PTB rate estimated using Garbhini-GA2 was more accurate when compared to the INTERGROWTH-21st and Hadlock formulae, which overestimated the rate by 22.47% and 58.91%, respectively.

*Corresponding author. Translational Health Science and Technology Institute, Faridabad, India.
**Corresponding author. Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India.
E-mail addresses: ramachandran@thsti.res.in (R. Thiruvengadam), sinha@iitm.ac.in (H. Sinha).
[j]These authors contributed equally and share joint first authorship.
[k]Members listed at the end of the paper.
[l]Senior author.

Interpretation The Garbhini-GA2 is the first late-trimester GA estimation model to be developed and validated using Indian population data. Its higher accuracy in GA estimation, comparable to GA estimation in the first trimester and PTB classification, underscores the significance of deploying population-specific GA formulae to enhance antenatal care.

Funding The GARBH-Ini cohort study was funded by the Department of Biotechnology, Government of India (BT/PR9983/MED/97/194/2013). The ultrasound repository was partly supported by the Grand Challenges India-All Children Thriving Program, Biotechnology Industry Research Assistance Council, Department of Biotechnology, Government of India (BIRAC/GCI/0115/03/14-ACT). The research reported in this publication was made possible by a grant (BT/kiData0394/06/18) from the Grand Challenges India at Biotechnology Industry Research Assistance Council (BIRAC), an operating division jointly supported by DBT-BMGF-BIRAC. The external validation study at CMC Vellore was partly supported by a grant (BT/kiData0394/06/18) from the Grand Challenges India at Biotechnology Industry Research Assistance Council (BIRAC), an operating division jointly supported by DBT-BMGF-BIRAC and by Exploratory Research Grant (SB/20-21/0602/BT/RBCX/008481) from Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), IIT Madras. An alum endowment from Prakash Arunachalam (BIO/18-19/304/ALUM/KARH) partly funded this study at the Centre for Integrative Biology and Systems Medicine, IIT Madras.

Keywords: Pregnancy dating; Gestational age; Second trimester; Third trimester; Preterm birth; GARBH-Ini cohort; Garbhini-GA2; Hadlock; INTERGROWTH-21st

Research in context

Evidence before this study

Accurate pregnancy dating is crucial for providing appropriate antenatal care and determining the delivery date. However, unlike GA estimation using crown-rump length in the first trimester, dating based on foetal biometry in the second and third trimesters is more susceptible to inaccuracies. This is a significant public health concern in LMICs like India, where about 30% of pregnant women seek initial antenatal care only after 14 weeks of gestation. The dating formulae currently used in LMICs have been developed using foetal biometry data from a Caucasian population, and using these formulae in ethnically diverse cohorts without appropriate modifications might lead to errors.

Added value of this study

This study developed a dating model, Garbhini-GA2, for the second and third trimesters of pregnancy based on several candidate biometric predictors measured in a population from North India. This model's performance was evaluated internally and validated in an external cohort. It outperformed the currently used dating models by reducing GA estimation median errors by more than three times with GA estimates comparable to the GA dating in the first trimester. Additionally, the Garbhini-GA2 model estimated PTB rates closer to the rate calculated in the first trimester, while the published formulae overestimated these PTB rates.

Implications of all the available evidence

While introducing the externally validated Garbhini-GA2 model, this study has clearly demonstrated that the most widely used Hadlock formula performed poorly in Indian settings. Garbhini-GA2 model, which uses routinely measured foetal biometry, can be swiftly adopted for clinical purposes across the Indian subcontinent after a pan-India validation. Applying the Garbhini-GA2 model will improve the clinical care of obstetricians and neonatologists, enhance the precision of epidemiological estimates for pregnancy outcomes, and provide accurately phenotyped participants for mechanistic research.

Introduction

Pregnancy dating or determining GA is crucial to obstetric care. Accurate pregnancy dating is essential for effective antenatal care, including investigations for foetal morphological anomalies, gestational diabetes, and preeclampsia. It becomes crucial during foetal growth monitoring and maternal nutritional supplementation. From an epidemiological perspective, it is also vital for obtaining reliable population-level estimates of pregnancy outcomes like stillbirth, preterm birth (PTB), and foetal growth restriction. Further, getting accurate GA will ensure correct phenotyping or classification of PTB to aid efficient clinical and biological research studies. Traditionally, Naegle's rule of counting days from the last menstrual period (LMP) has been used to estimate delivery dates and GA. However, this method depends on the regularity of menstrual cycles and accurate recall of the LMP date, which can be

affected by conditions like polycystic ovarian syndrome,[1] obesity,[2] contraceptive use, and preconception breastfeeding practices.[3,4] Due to its limitations, the use of LMP for dating is being replaced by ultrasound parameters. The ultrasound measurement of crown-rump length (CRL) in the first trimester is widely accepted as the most accurate method for estimating GA during pregnancy.[5–9] CRL measurement protocols are well-established and used as a standard of care worldwide.

However, according to the National Family Health Survey in India (NFHS-5) 2019–21, about 30% of pregnant women come for their first antenatal visit after the first trimester.[10] The dating methods for the late trimesters rely on foetal biometry. Established formulae like Hadlock[11] and INTERGROWTH-21st[12] utilise foetal biometry to estimate GA. Specifically, the Hadlock formula is commonly employed in India and globally as the pre-programmed formula in ultrasound machines to estimate GA based on CRL and foetal weight.[13,14] However, it is essential to note that this formula was developed using a small sample of pregnant Caucasian women from a North American population. Therefore, it may not be suitable for the South Asian population, where foetuses generally have smaller sizes.[15]

Since these formulae rely on foetal size, variations in foetal growth during the second and third trimesters can significantly affect the accuracy of GA estimation. This inaccuracy in the GA might affect the estimation of the due date of delivery leading to premature or delayed induction of labour and resultant complications.[16,17] It is evident that to improve the accuracy of GA estimation during pregnancy, it is crucial to identify biometric features less affected by foetal growth restriction, particularly in LMIC,[18] where the incidence of perinatal complications is high but resources are limited, or the pregnant women seek antenatal care very late for various reasons that include socio-economic and demographic factors.[19–23] Several attempts have been made to develop population-specific GA estimation models in LMICs.[24–26] Many of these attempts utilised the data from the Alliance for Maternal and Newborn Health Improvement (AMANHI) cohorts, which covered study sites in Bangladesh, Pakistan, and Tanzania.[27] These efforts implemented machine learning approaches and utilised data such as one-dimensional Doppler ultrasound, maternal blood pressure, birth weight, metabolite screening data, ultrasound blind sweep cineloop videos, etc. In this study, we aimed to develop dating models for the second and third trimesters of pregnancy using commonly measured biometric predictors that accurately estimate GA in late trimesters using the data from the GARBH-Ini (Interdisciplinary Group for Advanced Research on BirtH Outcomes DBT India Initiative) cohort, with the study site located in North India. The developed models were validated using an unseen dataset from the same cohort and an independent cohort from South India. Additionally, we evaluated the impact of our newly developed model on estimating the PTB rate and compared it to globally published models.

Methods

Study design and data collection

Site and participants—GARBH-Ini cohort

The GARBH-Ini cohort is an ongoing prospective hospital-based observational study of pregnant women initiated in May 2015 at Gurugram Civil Hospital, Gurugram, Haryana, India. This secondary care public-funded hospital delivers both primary and secondary-level antenatal care to pregnant women. In the GARBH-Ini cohort, we enrolled women who visited the antenatal clinic of the hospital and were willing to be enrolled and followed up in the cohort. The GARBH-Ini cohort is one of the largest studies on adverse birth outcomes, specifically preterm birth, in LMICs aiming to enrol 12,000 participants. For each participant, extensive data was collected–ultrasound maternal and foetal biometry and images, maternal anthropometric, clinical, obstetric and socioeconomic data and several biospecimens, including maternal blood samples at each visit, amniotic fluid, saliva, stool, high vaginal swab, placental tissue sample, and cord blood. The methods of the GARBH-Ini study have been previously published.[18] Briefly, the study enrolled participants before 20 weeks of gestation and monitored them thrice during pregnancy (at 18–20 weeks, 26–28 weeks, and 30–32 weeks) until delivery. A first trimester dating scan using CRL was performed if a participant was enrolled within 14 weeks of pregnancy. The participants were followed at least once each trimester until the end of their pregnancy. An ultrasound examination was performed each visit to assess the foetal biometry and other foetal and maternal characteristics. The ultrasound scans were performed in a standardised manner by qualified and certified radiologists using GE Voluson E8 Expert (General Electric Healthcare, Chicago, Illinois).

Eligible women were provided with a participant information sheet before participating in the study, and their written informed consent was obtained after they had read and understood the information. In the case of an eligible woman who was illiterate, the purpose of the study was explained to her, and her consent was confirmed through her thumb impression, provided she understood and explicitly expressed her consent verbally. In such instances, a literate, impartial witness signed the consent form on her behalf. The institutional ethics committees of Gurugram Civil Hospital, Safdarjung Hospital, Translational Health Science and Technology Institute, Christian Medical College, Vellore, and

Indian Institute of Technology Madras have approved the study.

The data for this study was derived from 6498 participants enrolled in the GARBH-Ini cohort. Amongst the enrolled participants, we included those who had their pregnancy dating conducted in the first trimester, underwent at least one ultrasound examination in the second or third trimester, and had documented pregnancy outcomes (N = 2649). The selection criteria and participant flow are presented in Fig. 1. Each observation for participants in the second and third trimesters was treated independently, resulting in a total sample size of 4972. Out of these observations, 4768 were chosen, corresponding to 2575 participants, based on the availability of the frequently used ultrasound sonography (USG) parameters for foetal measurements, which included biparietal diameter (BPD), occipito-frontal diameter (OFD), head circumference (HC), abdominal circumference (AC), and femur length (FL).

Site and participants—CMC Vellore cohort

We conducted an independent retrospective hospital-based cohort study at the CMC Vellore, Tamil Nadu, India, from July 2022 to January 2023, to validate the models built using the GARBH-Ini cohort data. We determined the *a priori* sample size of 922 for the external validation of dating models.[28] The detailed considerations for the sample size determination have been discussed in the Supplementary Methods. A total of 948 participants who had their pregnancy already

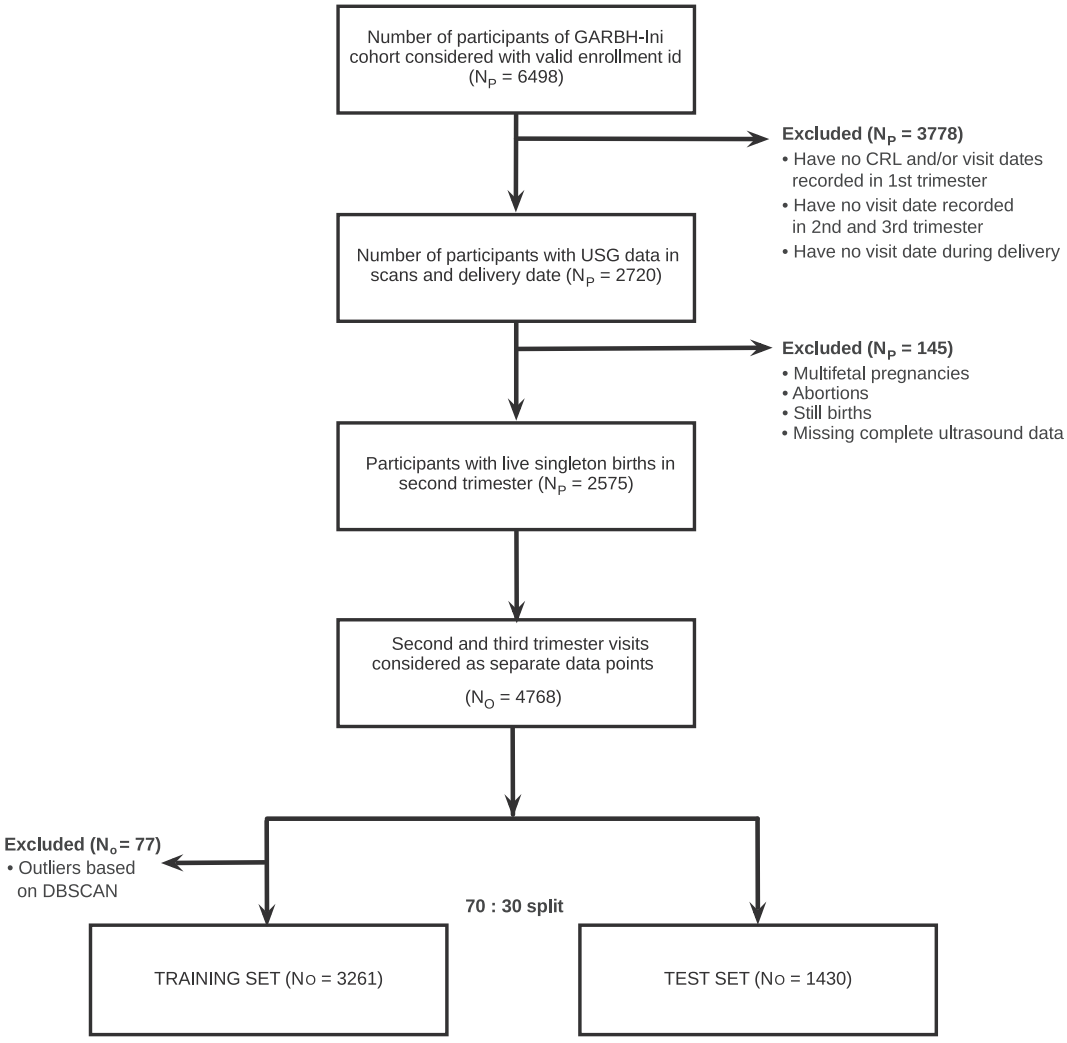

***Fig. 1:* Outline of the data selection process for TRAINING and TEST set**. Exclusion criteria for each step are indicated. N_p indicates the number of participants included or excluded by that criterion, and N_o shows the number of unique observations derived from the participants in a dataset.

dated in the first trimester using CRL measurement and visited the hospital for a follow-up scan in their second and third trimesters were included in the study. The ultrasound scans were performed on GE Voluson E10 (General Electric Healthcare, Chicago, Illinois) in a standardised manner by qualified and certified radiologists. The enrolment was done such that the scans of these participants were evenly distributed throughout the gestational period of 20–39 weeks.

Dataset preparation for modelling and validation

We divided the 4768 observations (N_o) in the GARBH-Ini dataset (complete-case analysis) into two sets: a TRAINING set containing 3338 observations (70% of the dataset) and an unseen TEST set containing 1430 observations (30% of the dataset, Fig. 1, Supplementary Methods). The participants in the TRAINING and TEST datasets were mutually exclusive. The TEST set was utilised for internal validation to evaluate the performance of published formulae and the models developed in this study. For the external validation of the models, we used 948 observations from the CMC Vellore cohort as the external VALIDATION set.

Definition of gold standard GA

In our previous study,[29] we developed a first trimester dating model known as the Garbhini-GA1, which utilised CRL measurement. In that study, we demonstrated that Garbhini-GA1 performed on par with the Hadlock and INTERGROWTH-21st first trimester dating models. In the present study, we utilised the Garbhini-GA1 model based on CRL measurements to estimate the first trimester GA of our participants. When participants returned for subsequent ultrasound scan visits in the second or third trimesters, we calculated the difference between the dates of their first and subsequent scans. We then added this time difference to the estimated first trimester GA to determine the expected GA at the second or third trimester visits. These calculated GA estimates were considered the "gold standard" or ground truth for developing our formulae for the second and third trimester models (Figure S1).

Feature selection

We conducted feature selection on 21 potential features (listed in Table S1) using our TRAINING set, which included the primary USG variables: BPD, OFD, HC, AC, and FL. We used the Boruta method based on a random forest classifier to perform the feature selection process. The Boruta algorithm created an extended dataset by first duplicating the dataset and shuffling the values in each column to create 'shadow features' that were combined with the original features. It then used a random forest classification algorithm to compare the significance of each feature by their Z-scores, retaining the features with scores higher than the maximum Z-score of its shadow features until all features were confirmed or rejected or the limit of random forest runs were reached (Supplementary Methods).

Development of population-specific gestational dating model

We employed multiple approaches to establish the relationship between foetal biometric variables and GA. These included Random Forest and Gradient Boosting machine learning methods implemented in R and multivariable polynomial regression developed using an optimisation technique inspired by the process of natural selection and genetics called "Genetic Algorithm" implemented in MATLAB (Supplementary Methods). We decided to implement these methods because they were the most appropriate for achieving our objectives, which involved identifying the most essential population-specific characteristics to determine GA accurately. Furthermore, these methods are well-suited for addressing the expected collinearity in the data, as we were working with variables associated with foetal growth, which were likely to be correlated and also capture if any non-linear relationships existed between these features. Random Forest and Gradient Boosting models were developed after undergoing hyperparameter tuning, which involved finding the optimal hyperparameter settings. This included parameters such as the number of decision trees and the maximum number of features considered for node splitting in the case of Random Forest and parameters like the number of boosting rounds, the maximum tree depth, and others for Gradient Boosting. The developed models were evaluated using the TEST and VALIDATION datasets (Supplementary Methods; Figure S3).

In the genetic algorithm method, we generated several candidate multivariate polynomial equations using the combinations of the USG variables in the TRAINING set, including their natural logarithms and square root forms (Table S3). Due to the stochastic nature of genetic algorithms, an optimal solution was not guaranteed. Therefore, the algorithm was run ten times independently to ensure reproducibility and consistency. We selected those equations that resolved in more than 5 out of 10 trials. These selected equations were evaluated using the TEST and VALIDATION datasets. The best-performing multivariate polynomial equation was named the Garbhini-GA2 formula.

Comparison of dating models

We evaluated the performance of the Garbhini-GA2 formula in predicting GA on both the TEST set (N_o = 1430) and the VALIDATION set (N_o = 948) by comparing it to the Hadlock formula,[11] $GA = 0.060(HP \times FL) + 0.67(BPD) + 0.168(AP) + 10.85$ and the INTERGROWTH-21st formula,[30] $ln(GA) = 0.03243(ln(HP))^2 + 0.001644(FL \times ln(HP)) + 3.813$, where GA was in weeks, HP, FL, AP, HP was in cm. HP was head perimeter and is the same as HC, and AP

was abdominal perimeter and is the same as AC. We assessed the error in weeks between the predicted and gold standard GA, ensuring the normality of the error distribution and the homogeneity of the variance assumptions before performing statistical tests. To compare the distributions, we used the Kruskal–Wallis test followed by Dunn's posthoc test to identify the pairwise differences between formulae, which were corrected for multiple testing comparisons using the Bonferroni correction method. We also created parity plots for each of the three formulae between predicted and gold standard GA. Finally, we conducted a Bland-Altman analysis to evaluate the bias between the models and reported the pairwise mean difference and limits of agreement. All statistical analyses were performed using R packages (details in Supplementary Methods).

Preterm birth (PTB) analysis

To determine the prevalence of PTB in the TEST set, we computed the number of participants with a predicted GA of less than 37 weeks per 100 participants after adding the difference between the GA at birth (collected in the GARBH-Ini cohort) and the gold standard GA for three formulae–Hadlock, INTERGROWTH-21st, and Garbhini-GA2. To estimate the confidence interval of each prediction, we employed the Clopper-Pearson method. We used the Jaccard similarity coefficient to assess the agreement between the gold standard and the three formulae for preterm labelling. To visualise the classification of participants as preterm or not, we generated a quadrant plot, plotting the predicted GA on the x-axis against the gold standard GA on the y-axis.

Ethics approval and consent to participate

Ethics approvals were obtained from the Institutional Ethics Committees of Translational Health Science and Technology Institute; District Civil Hospital, Gurugram; Safdarjung Hospital, New Delhi (ETHICS/GHG/2014/1.43); Indian Institute of Technology Madras (IEC/2019-03/HS/01/07); and Christian Medical College Vellore (IRB Min. No. 14636 [INTERVEN]). Written informed consent was obtained from all study participants enrolled in the GARBH-Ini cohort. All the methods were performed following the relevant guidelines and regulations.

Role of the funding source

The study sponsors did not play a part in the design, data collection, analysis, interpretation, or manuscript drafting.

Results

Description of participants included in the study

GARBH-Ini cohort

The study recruited participants from the GARBH-Ini cohort, whose data was utilised for both the TRAINING and TEST sets. The median age of the participants was 23 years. The median weight and height in the TRAINING set were 56.5 kg (interquartile range: 51.0–63.4) and 153.1 cm (interquartile range: 149.4–157), respectively. Most participants (57.2%) had a normal BMI, with a median of 20.4 (interquartile range: 18.2–23.1). More than half of the participants (51.0%) were primigravida, and most belonged to middle or lower socioeconomic strata as per Modified Kuppuswamy scale.[31] The median GA at recruitment was 19.4 weeks (interquartile range: 19.1–20.1). Additional baseline characteristics of the participants are presented in Table 1.

CMC Vellore cohort

The data were collected from the CMC Vellore cohort for the external VALIDATION set. The median age of the participants in this cohort was 27 years, and the median GA in the first trimester was 10 weeks and 3 days.

Feature selection

After analysing the importance of features using the Boruta algorithm, we determined that 9 out of 22 features in the GARBH-Ini cohort dataset significantly predicted GA (Figure S2). These features include BPD, OFD, FL, HC, AC, symphysiofundal height, BMI, maternal weight, and abdominal girth. Notably, the top 5 features among these nine were USG-based metrics, namely AC, FL, OFD, HC, and BPD, which displayed a significant difference in importance compared to non-USG metrics. These five features are routinely recorded in clinics across India, making them ideal for developing an Indian population-specific model that can be readily applied in clinics throughout the country. Further, these five features also stood out in the feature selection analysis (Figure S2). Consequently, only these USG-based metrics were used to develop this study's GA models. Given this finding, we primarily focused on collecting the USG-based metrics in our CMC Vellore cohort, which we used as the external dataset to validate the developed models.

Garbhini-GA2 formula

Using the five USG-metric-based variables and their transformed versions from the TRAINING set as inputs in the Genetic Algorithm, we short-listed 18 multivariate polynomial formulae in ten trials of the algorithm (Table S4). The performance evaluation of these 18 formulae and the models developed using Random Forest and Gradient Boosting methods indicated that the Random Forest and Gradient Boosting models were particularly overfitting and performed worse in the VALIDATION set with RMSE values of 2.02 and 2.33, and R^2 values of 0.87 and 0.84, respectively. This result was in contrast to their performance in the TRAINING set, where the RMSE values were 0.91 and 0.90, and R2 values were 0.98 for both models (Figure S3 and S4 and Table S5 and S6).

Upon ranking the formulae generated by the Genetic Algorithm, based on the number of times the formula was resolved and the achieved RMSE in the TRAINING set, we observed that the top two formulae contained the AC variable (Table S4). However, AC is typically difficult to measure accurately from ultrasound images due to poor foetal abdomen, contrast against the surroundings, non-uniform contrast, and irregular shape.[32–34] Additionally, AC is also most affected by foetal growth restriction[35]; hence we selected the next best formula, which did not include AC from the sorted list, $ln(GA) = 0.09255(ln(HC)^2) + 0.07661(ln(BPD) \times ln(OFD)) + 2.05685$ as the Garbhini-GA2 formula, where GA was in weeks, HC, BPD, OFD were in cm. The Garbhini-GA2 was resolved in all ten trials of the algorithm (Table S4), outperforming both the Random Forest and Gradient Boosting models in the VALIDATION set with an RMSE of 1.41 and an R^2 of 0.92. Moreover, its performance in the TEST set revealed an RMSE value of 1.17 and an R^2 of 0.97, affirming its consistency when compared to the Random Forest and Gradient Boosting methods in both the TEST and VALIDATION sets. Considering the fact that some of the participants may have more than one measurement, which might be correlated, we randomly selected one of the observations of each participant and repeated genetic algorithm model building. The model from this random set was similar to the Garbhini-GA2 formula (Table S7).

Comparison of Garbhini-GA2 and published formulae in the second and third trimesters

The Garbhini-GA2 was found to have the lowest median error in predicting GA compared to the Hadlock and INTERGROWTH-21st models in both the TEST and VALIDATION sets, as determined by the comparison of error distributions (Table S8). This error distribution difference was statistically significant (p-value <0.05; Fig. 2A and B). The Bland-Altman analysis for bias between the formulae and gold standard GA also showed that Garbhini-GA2 predictions had a minor mean difference of around half a day (TEST set: –0.079 weeks; 95% CI: –2.359, 2.201; VALIDATION set: –0.068 weeks; 95% CI: –2.837, 2.7). In comparison, the INTERGROWTH-21st and Hadlock formulae showed a mean difference of about 1.5–3 days and 5–7 days in TEST and VALIDATION sets, respectively. This indicated that Garbhini-GA2 had a high level of agreement with the gold standard GA (Table 2). This agreement can be observed in the parity plots (Figure S6 and S7), where the majority of data points for Garbhini-GA2 and INTERGROWTH-21st were located on or close to the straight line, while Hadlock distribution showed higher disagreement.

These findings suggested that the Garbhini-GA2 estimates of GA were comparable to GA estimated in the first trimester and, thus, was a more accurate formula for predicting GA, particularly during the later trimesters.

Sociodemographic characteristics	TRAINING set median (IQR) or N (%) or Mean ± SD	TEST set median (IQR) or N (%) or Mean ± SD
Age (years)	23 (21,26)	23 (21,26)
GA at enrolment by USG (weeks)	20.9 ± 4.2	20.9 ± 4.2
BMI at enrolment into the cohort[a]		
Underweight	27.8%	24.4%
Normal	57.2%	61.0%
Obese	12.0%	11.0%
Overweight	1.9%	2.3%
Haemoglobin (g/dL)	9 (8.5–9.5)	9 (8.4–9.5)
Weight (kgs)	56.5 (51.0–63.4)	57.1 (51.5–63.4)
Height (cm)	153.1 (149.4–157)	152.7 (149.2–156.5)
Socioeconomic status[b]		
0	0.4%	0.6%
1	18.1%	19.3%
2	37.9%	33.4%
3	42.8%	45.6%
4	0.2%	0.5%
Undetermined	0.5%	0.5%
Parity (number)		
0	51.0%	50.2%
1	33.6%	32.4%
2	12.1%	13.5%
3	2.6%	3.2%
4	0.7%	0.6%
5	0.1%	0
6	0	0
7	0	0
Level of education		
Illiterate	18.2%	17.8%
Literate or primary school	9.6%	12.7%
Middle school	16.0%	13.9%
High school	22.5%	22.4%
Post-high school diploma	17.1%	17.5%
Graduate	13.1%	13.5%
Post-graduate	2.9%	1.8%
Occupation		
Unemployed	92.8%	91.4%
Unskilled worker	3.2%	4.8%
Semi-skilled worker	1.4%	1.7%
Skilled worker	1.9%	1.2%
Clerk, shop, farm owner	0.1%	0.1%
Semi-professional	0.3%	0.1%
Professional	0.2%	0.6%
Fuel used for cooking[c]		
Biomass fuel	93.0%	92.5%
Clean fuel[d]	7.0%	7.5%
Source of drinking water		
Safe water[e]	56.6%	58.8%
Unsafe water	43.4%	41.2%
Second-hand tobacco smoke		
Exposed	19.4%	18.5%
Unexposed	0.1%	81.4%
Undetermined	80.5%	0.1%

(Table 1 continues on next page)

Sociodemographic characteristics	TRAINING set median (IQR) or N (%) or Mean ± SD	TEST set median (IQR) or N (%) or Mean ± SD
(Continued from previous page)		
History of any chronic illnesses[f]		
Absent	98.3%	97.5%
Present	1.7%	2.5%
History of hypertensive disease of pregnancy		
Absent	98.9%	99.1%
Present	1.1%	0.9%
History of contraceptives at the time of conception		
Absent	95.9%	95.1%
Present	4.1%	4.9%

[a]Pre-pregnancy BMI was calculated as weight (kg)/height 2 (m) from participants' weight and height measured at enrolment. BMI categories were defined as underweight (<18.5); normal (18.5–24.9); overweight (25.0–29.9); obese (≥30.0). [b]Socioeconomic status was assessed using Modified Kuppuswamy's socioeconomic scale, calculated using the education and occupation of the head of the family and monthly family income. [c]Indoor air pollution: use of biomass fuel for cooking or the presence of a smoker in the residential compound, as reported by the participant. [d]Clean fuel includes liquefied petroleum gas and electricity. [e]Safe water includes bottled water or piped water into the residence. [f]Chronic illnesses include a history of hypertension, diabetes, cardiac disease and thyroid disorders.

Table 1: **Baseline characteristics of the participants included in the TRAINING (N_p = 1803, N_o = 3338, before outlier removal) and TEST (N_p = 772, N_o = 1430) sets.**

Impact of choice of dating formula on the estimation of preterm rates

When the gold standard for GA was used, the PTB rate in the TEST set was 10% (CI 8.49, 11.67). However, when different models were used for pregnancy dating, the estimated PTB rates on the TEST set varied between 11.89% and 21.82%. Among all the models, Garbhini-GA2 (11.89%; CI 10.26, 13.68) yielded PTB rates closest to those estimated by the gold standard based on first trimester CRL. Notably, INTERGROWTH-21st (14.90%; CI 13.09, 16.85) and Hadlock (21.82%; CI 19.70, 24.05) formulae overestimated PTB rates (Fig. 3A). We found the Garbhini-GA2 as the most accurate, with 95.03% accuracy in identifying the preterm and term births correctly, followed by the INTERGROWTH-21st (93.42%) and Hadlock (87.2%, Fig. 3B–D).

We used the Jaccard similarity coefficient to compare the accuracy of three formulae in estimating GA at birth and identifying preterm births. This coefficient measured the overlap between the preterm outcomes based on each formula's GA estimates at birth and the gold standard dating method, where higher values indicated greater agreement. Our analysis revealed that the Garbhini-GA2 model had the highest Jaccard similarity coefficient (63.02%), while the Hadlock formula had the lowest (42.63%). This indicated that the Garbhini-GA2 method had the least misclassified preterm births among the published formulae we tested. The summary of these findings is presented in Table S8.

Discussion

In this study, we successfully developed the Garbhini-GA2 model to estimate GA during the second and third trimesters of pregnancy. Our Garbhini-GA2 estimations were closest to the gold-standard GA, determined using our Garbhini-GA1 model based on CRL measurements. In a prior study, we introduced the Garbhini-GA1 formula,[29] specifically designed for the Indian population, showing comparable performance to Hadlock and INTERGROWTH-21st formulae but notably higher sensitivity in estimating preterm birth rates. We utilised Garbhini-GA1 to establish the gold standard GA for later trimesters in this study, facilitating a comprehensive performance comparison among Garbhini-GA2, Hadlock, and INTERGROWTH-21st.

From an epidemiological perspective, combining our previously published Garbhini-GA1[29] model with the newly introduced Garbhini-GA2 dating formula enhances the precision of pregnancy outcome estimates, including preterm birth, small for gestational age (SGA), and stillbirth in the Indian population, where accurate GA is crucial. Precise GA estimation is also vital for managing extreme[36] and moderate[37] preterm births in obstetric and neonatal care, as postnatal care frameworks rely on it. Beyond the medical field, Indian population-specific dating models will support biologists in refining the clinical phenotyping of birth outcomes for biomarker and mechanistic studies.

We demonstrated that the Garbhini-GA2 performed with a lower median error in GA estimation than the widely accepted Hadlock formula. The improved performance of the Garbhini-GA2 model, as compared with Hadlock, indicated the variation in foetal biometry between the Caucasian and Indian populations. Additionally, the PTB rate estimated by Garbhini-GA2 was closely compared to that calculated from the gold standard first-trimester dating model, with only an 18.90% overestimation. In contrast, PTB rate estimates using INTERGROWTH-21st and Hadlock formulae were 22.47% and 58.91% higher, respectively. Notably, the overestimation of the PTB rate by the INTERGROWTH-21st and Hadlock formulae, compared to the Garbhini-GA2 model, demonstrated closer to true epidemiological estimates of PTB rates by the latter, specific for the region.

The Garbhini-GA2 model presented here was chosen by thoroughly evaluating the performance of all polynomial regression equations we generated, considering potential challenges associated with the features included in the equation. As was the case with AC, identifying the right plane for the measurement of FL was perceived to be a challenge by the clinicians. Therefore, incorporating AC and FL into the formula might introduce a variable measurement, potentially resulting in dating errors.

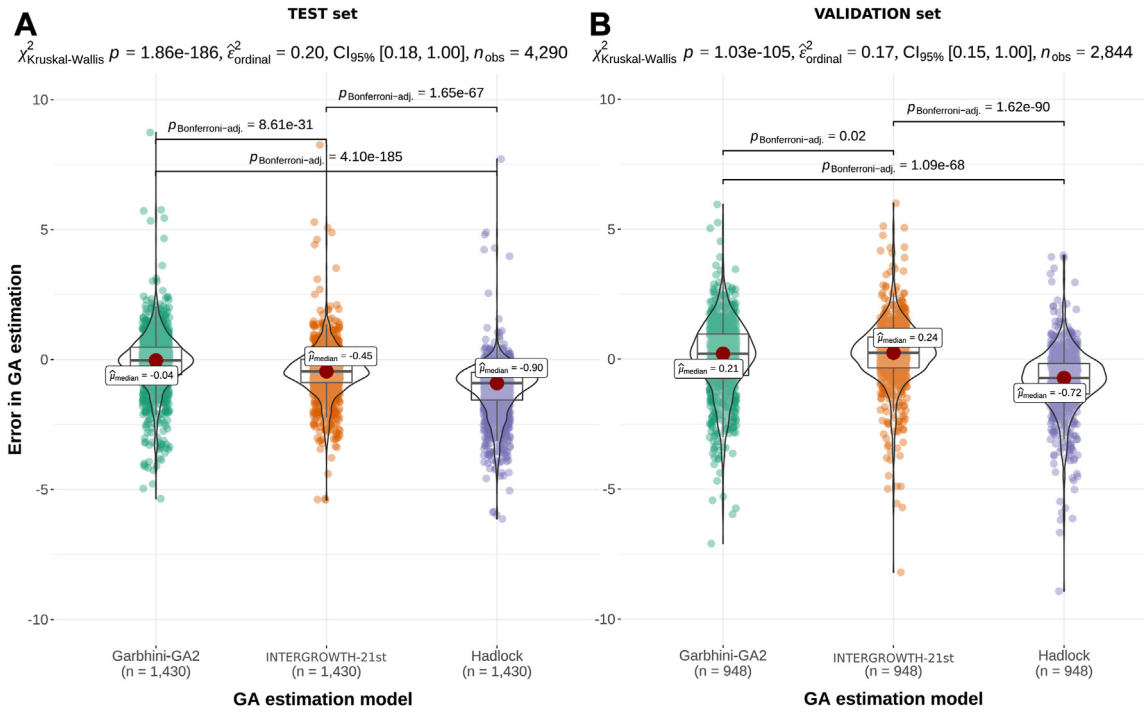

Fig. 2: **Performance of Garbhini-GA2 model for GA estimation in (A) TEST and (B) VALIDATION sets.** Distribution plots represent the error (in weeks) in the estimation of GA compared to the gold standard GA. The y-axis is the error (the difference between model-predicted and gold standard GA), and the x-axis is the density for each model in the (A) TEST set (N_o = 1430) and (B) VALIDATION set (N_o = 948). Due to the ordinal distribution observed in our data, we utilised the nonparametric Kruskal–Wallis (KW) test to compare the error distribution among the different formulae. The p-value obtained from the KW test ($\chi^2_{Kruskal-Wallis}$) is presented in the top left corner of the plots. The partial epsilon-squared value ($\hat{\varepsilon}^2_{ordinal}$), along with its 95% confidence intervals ($CI_{95\%}$), is provided alongside the KW p-value to indicate the effect size (ES) of the KW test. Additionally, the total sum of all observations (n_{obs}) is displayed next to the ES value. The corrected p-values resulting from the Dunn test ($p_{Bonferroni\text{-}adj.}$) for significant KW are shown on top of the error distributions, enabling pairwise comparisons between the formulae.

Furthermore, given that the Garbhini-GA2 model was developed using data from one cohort and subsequently validated on an external, independent cohort that closely mirrors real-world clinical conditions, it offered significant advantages. This approach enhanced the generalisability and robustness of the model by ensuring its performance was not confined to a specific dataset but could be extrapolated to different populations. An additional advantage of Garbhini-GA2 is its simplicity as a polynomial regression equation, making it straightforward to understand and integrate into existing clinical workflows for accurate GA prediction. It offers transparency in its structure, enhancing interpretability for clinicians and healthcare professionals, which not only aids in the seamless adoption of the model but also facilitates training and education for healthcare staff.

While our proposed Garbhini-GA2 model offers notable advantages, we also acknowledge a limitation in our study. The diversity of the population used for the development and validation of Garbhini-GA2 may only capture a subset of the overall demographic and ethnic

	Gold standard	Hadlock	INTERGROWTH-21st	Garbhini-GA2
Gold standard		−1.089 (−3.302, 1.125)	−0.447 (−2.554, 1.66)	−0.079 (−2.359, 2.201)
Hadlock	0.824 (−1.504, 3.152)		0.642 (−0.315, 1.599)	1.01 (−0.203, 2.222)
INTERGROWTH-21st	−0.226 (−2.602, 2.151)	−1.049 (−2.002, −0.096)		0.368 (−0.892, 1.628)
Garbhini-GA2	−0.068 (−2.837, 2.7)	−0.892 (−2.214, 0.43)	0.157 (−1.765, 2.079)	

The values in the table represent the pairwise mean difference (in weeks) between the formulae and the gold standard GA in the VALIDATION and TEST sets derived from the CMC Vellore and GARBH-Ini cohorts, respectively. The top diagonal values correspond to the TEST set, and the bottom values correspond to the VALIDATION set.

Table 2: **Bland Altman analysis and PTB agreement on TEST set.**

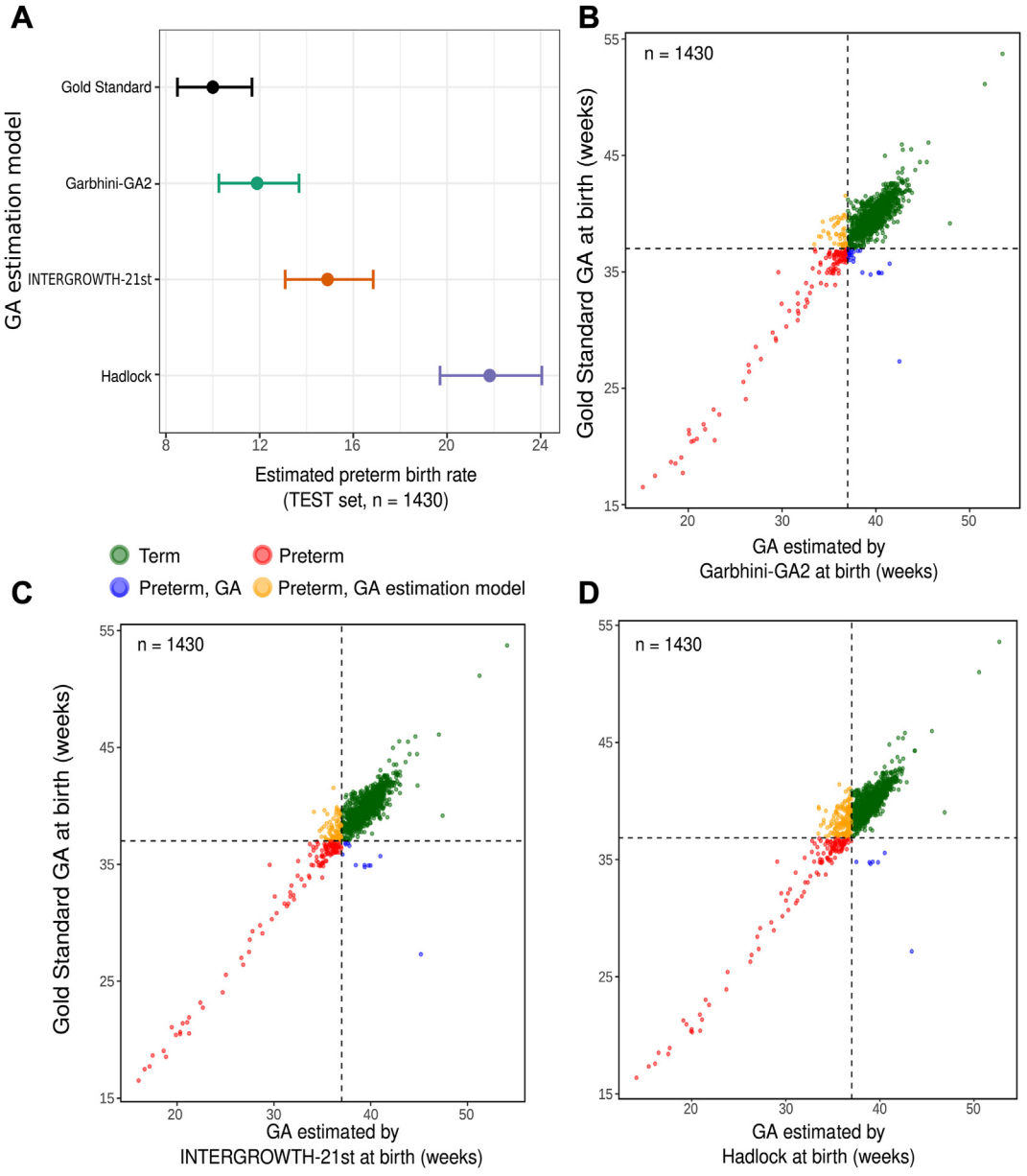

***Fig. 3*: Performance of Garbhini-GA2 model in terms of PTB rate in TEST set derived from the GARBH-Ini cohort**. (A) PTB rates by various models: PTB rates are labelled by each model with 95% confidence intervals on the TEST set (N_o = 1430). (B–D) Comparison of individual-level classification of preterm birth by a model and gold standard GA. Green (term birth for both), red (preterm birth for both), blue (term birth for gold standard but preterm birth for model) and purple (term for model but preterm for gold standard).

landscape in India. Further, the CMC Vellore retrospective hospital-based cohort was not as well characterised as the GARBH-Ini cohort, which hampers a comprehensive understanding of the performance of our model in an external setting. To overcome these limitations, we are committed to an extensive validation process. Our validation strategy will involve enrolling pregnant women from various demographic and ethnic backgrounds, including but not limited to different geographical regions, socioeconomic status, and cultural groups. This multifaceted approach will enhance the generalisability of the Garbhini-GA2 model.

We have developed the Garbhini-GA2 model in this study, which offers an accurate and easily applicable way to estimate GA during the second and third trimesters of pregnancy. This model aims to overcome the challenges faced in LMICs like India, where pregnant women often miss the opportunity for accurate GA

dating as they seek antenatal care after the first trimester. With further comprehensive validation, the Garbhini-GA2 model has the potential to be rapidly adopted for clinical use across Southeast Asian regions. Its application will benefit clinicians, epidemiologists, and biologists in improving pregnancy outcome estimates and clinical phenotyping of birth outcomes.

Contributors

SB, RT, and HS conceived this study; VPG, ND, AX, SBT, and RV performed data and statistical analyses; BKD and RT performed data exports and contributed to data analysis; SM, NW, AK, RT, and GARBH-Ini Study Group developed and implemented the clinical data collection methods and data management in the GARBH-Ini cohort; SR, AA, SBE, AGC developed and implemented the clinical data collection methods and data management at CMC Vellore, RR provided critical feedback on data analysis; VPG, ND, AX, SBT, RV, BKD, RR, SB, RT, and HS interpreted the results; VPG, RT, and HS verified the data, compiled the results; VPG, ND, AX, RT, and HS wrote the first draft of the manuscript and all listed authors critically revised and edited subsequent manuscript drafts; all authors had full access to the data and reviewed and approved the manuscript.

Data sharing statement

The datasets used and analysed in the current study are available upon reasonable request from the corresponding author after approval of the DBT Steering Committee of GARBH-Ini. All the codes used for this paper are available at https://github.com/HimanshuLab/Garbhini-GA2.

Declaration of interests

The authors declare the following competing interests. HS, RT, and SB are inventors of the Garbhini-GA2 formula (Indian Patent Application no: 202341061748), submitted by the Indian Institute of Technology Madras.

Acknowledgements

We thank all the participants of the GARBH-Ini study and CMC Vellore. We thank members of the Centre for Integrative Biology and Systems Medicine (IBSE), Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), IIT Madras, and Aryabhata Data Science and AI Program at THSTI (ADAPT), THSTI. We thank H. Sabiha Fathima and Vallalarasi A. from CMC Vellore for the data collection. We also thank Balaraman Ravindran, IIT Madras, India; Gagandeep Kang, Bill and Melinda Gates Foundation, USA; and Ashok Venkitaraman, Cancer Science Institute of Singapore, Singapore, for their valuable suggestions. M.K. Bhan will always be remembered reverently for his critical scientific and technical feedback. We sincerely appreciate the Department of Biotechnology, Government of India, for supporting the GARBH–Ini program. We would like to express our gratitude to Arshi Mehboob and Shirshendu Mukherjee, who provided support for this research through Grand Challenges India at the Biotechnology Industry Research Assistance Council of the Government of India. Finally, we recognise the efforts of the research physicians, study nurses, clinical and laboratory technicians, field workers, the internal quality improvement team, and the project and data management team and all the members of the GARBH–Ini cohort.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lansea.2024.100362.

References

1 Lobo RA. What are the key features of importance in polycystic ovary syndrome? *Fertil Steril*. 2003;80(2):259–261. https://doi.org/10.1016/s0015-0282(03)00733-7.
2 Wei S, Schmidt MD, Dwyer T, Norman RJ, Venn AJ. Obesity and mensural irregularity: associations with SHBG, testosterone, and insulin. *Obesity*. 2009;17:1070–1076. https://doi.org/10.1038/oby.2008.641.
3 Chowdhury R, Sinha B, Sankar MJ, et al. Breastfeeding and maternal health outcomes: a systematic review and meta-analysis. *Acta Paediatr*. 2015;104(467):96–113. https://doi.org/10.1111/apa.13102.
4 Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. *Contraception*. 2004;70(4):289–292. https://doi.org/10.1016/j.contraception.2004.04.012.
5 Committee on obstetric practice, the American Institute of ultrasound in medicine, and the society for maternal-fetal medicine. Committee opinion no 700: methods for estimating the due date. *Obstet Gynecol*. 2017;129(5):e150–e154. https://doi.org/10.1097/AOG.0000000000002046.
6 Geirsson RT. Ultrasound instead of last menstrual period as the basis of gestational age assignment. *Ultrasound Obstet Gynecol*. 1991;1(3):212–219. https://doi.org/10.1046/j.1469-0705.1991.01030212.x.
7 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(6):587–596. https://doi.org/10.1111/j.1365-3016.2008.00965.x.
8 Bennett KA, Crane JMG, O'shea P, Lacelle J, Hutchens D, Copel JA. First trimester ultrasound screening is effective in reducing postterm labor induction rates: a randomized controlled trial. *Am J Obstet Gynecol*. 2004;190(4):1077–1081. https://doi.org/10.1016/j.ajog.2003.09.065.
9 Sarris I, Ioannou C, Chamberlain P, et al. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol*. 2012;39(3):266–273. https://doi.org/10.1002/uog.10082.
10 National family health survey (NFHS-5) India report | international Institute for population sciences (IIPS). https://iipsindia.ac.in/content/national-family-health-survey-nfhs-5-india-report. Accessed June 21, 2023.
11 Hadlock FP, Deter RL, Harrist RB, Park SK. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. *Radiology*. 1984;152(2):497–501. https://doi.org/10.1148/radiology.152.2.6739822.
12 Papageorghiou AT, Kennedy SH, Salomon LJ, et al. The INTERGROWTH-21st fetal growth standards: toward the global integration of pregnancy and pediatric care. *Am J Obstet Gynecol*. 2018;218(2S):S630–S640. https://doi.org/10.1016/j.ajog.2018.01.011.
13 Aggarwal N, Sharma GL. Fetal ultrasound parameters: reference values for a local perspective. *Indian J Radiol Imaging*. 2020;30(2):149–155. https://doi.org/10.4103/ijri.IJRI_287_19.
14 Warshafsky C, Ronzoni S, Quaglietta P, et al. Accuracy of Hadlock IV and fetal weight estimation in preterm premature rupture of membranes. *J Obstet Gynaecol Can*. 2020;42(5):694. https://doi.org/10.1016/j.jogc.2020.02.107.
15 Kierans WJ, Joseph KS, Luo ZC, Platt R, Wilkins R, Kramer MS. Does one size fit all? The case for ethnic-specific standards of fetal growth. *BMC Pregnancy Childbirth*. 2008;8(1):1. https://doi.org/10.1186/1471-2393-8-1.
16 Källén K. Increased risk of perinatal/neonatal death in infants who were smaller than expected at ultrasound fetometry in early pregnancy. *Ultrasound Obstet Gynecol*. 2004;24(1):30–34. https://doi.org/10.1002/uog.1082.
17 Thorsell M, Kaijser M, Almström H, Andolf E. Expected day of delivery from ultrasound dating versus last menstrual period–obstetric outcome when dates mismatch. *BJOG*. 2008;115(5):585–589. https://doi.org/10.1111/j.1471-0528.2008.01678.x.
18 Bhatnagar S, Majumder PP, Salunke DM. Interdisciplinary group for advanced research on birth outcomes-DBT India initiative (GARBH-Ini). A pregnancy cohort to study multidimensional correlates of preterm birth in India: study design, implementation, and baseline characteristics of the participants. *Am J Epidemiol*. 2019;188:621–631.
19 Myer L, Harrison A. Why do women seek antenatal care late? Perspectives from rural South Africa. *J Midwifery Womens Health*. 2003;48(4):268–272. https://doi.org/10.1016/s1526-9523(02)00421-x.
20 Finlayson K, Downe S. Why do women not use antenatal services in low- and middle-income countries? A meta-synthesis of qualitative studies. *PLoS Med*. 2013;10(1):e1001373. https://doi.org/10.1371/journal.pmed.1001373.
21 Chimatiro CS, Hajison P, Chipeta E, Muula AS. Understanding barriers preventing pregnant women from starting antenatal clinic in the first trimester of pregnancy in Ntcheu District-Malawi.

Reprod Health. 2018;15(1):158. https://doi.org/10.1186/s12978-018-0605-5.
22 Ewunetie AA, Munea AM, Meselu BT, Simeneh MM, Meteku BT. DELAY on first antenatal care visit and its associated factors among pregnant women in public health facilities of Debre Markos town, North West Ethiopia. *BMC Pregnancy Childbirth*. 2018;18(1):173. https://doi.org/10.1186/s12884-018-1748-7.
23 Tripathy A, Mishra PS. Inequality in time to first antenatal care visits and its predictors among pregnant women in India: an evidence from national family health survey. *Sci Rep*. 2023;13(1):4706. https://doi.org/10.1038/s41598-023-31902-3.
24 Valderrama CE, Marzbanrad F, Hall-Clifford R, Rohloff P, Clifford GD. A proxy for detecting IUGR based on gestational age estimation in a Guatemalan rural population. *Front Artif Intell*. 2020;3:56. https://doi.org/10.3389/frai.2020.00056.
25 Gomes RG, Vwalika B, Lee C, et al. A mobile-optimized artificial intelligence system for gestational age and fetal malpresentation assessment. *Commun Med*. 2022;2(1):128. https://doi.org/10.1038/s43856-022-00194-5.
26 Pokaprakarn T, Prieto JC, Price JT, et al. AI estimation of gestational age from blind ultrasound sweeps in low-resource settings. *NEJM Evid*. 2022;1(5). https://doi.org/10.1056/evidoa2100058.
27 Aftab F, Ali AS. Cohort profile: the alliance for maternal and newborn health improvement (AMANHI) biobanking study. *Int J Epidemiol*. 2022;50:1780–1781.
28 Archer L, Snell KIE, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Stat Med*. 2021;40(1):133–146. https://doi.org/10.1002/sim.8766.
29 Vijayram R, Damaraju N, Xavier A, et al. Comparison of first trimester dating methods for gestational age estimation and their implication on preterm birth classification in a North Indian cohort. *BMC Pregnancy Childbirth*. 2021;21(1):343. https://doi.org/10.1186/s12884-021-03807-4.
30 Papageorghiou AT, Kemp B, Stones W, et al. Ultrasound-based gestational-age estimation in late pregnancy. *Ultrasound Obstet Gynecol*. 2016;48(6):719–726. https://doi.org/10.1002/uog.15894.
31 Wani RT. Socioeconomic status scales-modified Kuppuswamy and Udai Pareekh's scale updated for 2019. *J Family Med Prim Care*. 2019;8(6):1846–1849. https://doi.org/10.4103/jfmpc.jfmpc_288_19.
32 Skinner C, Mount CA. *Sonography assessment of gestational age*. StatPearls Publishing; 2023. https://www.ncbi.nlm.nih.gov/books/NBK570610/. Accessed November 26, 2023.
33 Yu J, Wang Y, Chen P, Shen Y. Fetal abdominal contour extraction and measurement in ultrasound images. *Ultrasound Med Biol*. 2008;34(2):169–182. https://doi.org/10.1016/j.ultrasmedbio.2007.06.026.
34 Jang J, Park Y, Kim B, Lee SM, Kwon JY, Seo JK. Automatic estimation of fetal abdominal circumference from ultrasound images. *IEEE J Biomed Health Inform*. 2018;22(5):1512–1520. https://doi.org/10.1109/jbhi.2017.2776116.
35 Long M, Nakahara A, Elmayan A, Tivis R, Biggio J, Williams F. Fetal growth restriction defined by abdominal circumference alone predicts perinatal mortality. *Am J Obstet Gynecol*. 2022;226(1):S179. https://doi.org/10.1016/j.ajog.2021.11.312.
36 Se HM. Perinatal management of extreme preterm birth before 27 weeks of gestation: a framework for practice. *Arch Dis Child Fetal Neonatal Ed*. 2020;105:232–239. https://doi.org/10.1136/archdischild-2019-318402.
37 Framework: early postnatal care of the moderate-late preterm infant. British Association of Perinatal Medicine. https://www.bapm.org/resources/framework-early-postnatal-care-of-the-moderate-late-preterm-infant. Accessed November 26, 2023.