Ultrasound based gestational age estimation in late pregnancy

Aris T. Papageorghiou[a] , Bryn Kemp[a], William Stones[b,c], Eric O. Ohuma[a,d], Stephen H. Kennedy[a], Manorama Purwar[e], Laurent J. Salomon[f], Douglas G. Altman[d], Julia A Noble[g], Enrico Bertino[h], Michael G. Gravett[i], Ruyan Pang[j], Leila Cheikh Ismail[a], Fernando C. Barros[k,l], Ann Lambert[a], Yasmin A. Jaffer[m], Cesar G. Victora[l], Zulfiqar A. Bhutta[n,o] and José Villar[a] for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st).

a. Nuffield Department of Obstetrics & Gynaecology and Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK
b. Faculty of Health Sciences, Aga Khan University, Nairobi, Kenya
c. School of Medicine, University of St Andrews, St Andrews, Scotland
d. Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Oxford, UK
e. Nagpur INTERGROWTH-21st Research Centre, Ketkar Hospital, Nagpur, India
f. Maternité Necker-Enfants Malades, AP-HP, Université Paris Descartes, Paris, France
g. Department of Engineering Science, University of Oxford, Oxford, UK
h. Dipartimento di Scienze Pediatriche e dell'Adolescenza, Cattedra di Neonatologia, Università degli Studi di Torino, Torino, Italy
i. Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), Seattle, WA,USA
j. School of Public Health, Peking University, Beijing, China
k. Programa de Pós-Graduação em Saúde e Comportamento, Universidade Católica de Pelotas, Pelotas, RS, Brazil
l. Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil
m. Department of Family & Community Health, Ministry of Health, Muscat, Sultanate of Oman
n. Division of Women & Child Health, The Aga Khan University, Karachi, Pakistan
o. Center for Global Health, Hospital for Sick Children, Toronto, ON, Canada

Address correspondence to: Prof Aris Papageorghiou, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Women's Centre, Level 3, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK; e-mail: aris.papageorghiou@obs-gyn.ox.ac.uk

Short title: International late pregnancy dating

Key words: gestational age; dating; fetal growth;

ABSTRACT

Objectives: Accurate gestational age (GA) estimation, preferably by ultrasound measurement of fetal crown rump length before 14 weeks' gestation, is an important component of high-quality antenatal care. The objective was to determine how GA can best be estimated by fetal ultrasound for women who present for the first time late in pregnancy with uncertain or unknown menstrual dates.

Methods: INTERGROWTH-21st was a large, prospective, multicentre, population-based project performed in eight geographically defined urban populations. One of its principal components, the Fetal Growth Longitudinal Study (FGLS), aimed to develop international fetal growth standards. Each participant had certain menstrual dates confirmed by first trimester ultrasound. Fetal head circumference (HC), biparietal diameter (BPD), occipitofrontal diameter (OFD), abdominal circumference (AC) and femur length (FL) were measured every 5 weeks from 14 weeks' gestation until birth. For each participant, a single, randomly selected ultrasound examination was used to explore all candidate biometric variables and permutations to build models to predict GA. Regression equations were ranked based upon minimization of the mean prediction error, goodness of fit and model complexity. An automated machine learning algorithm, the Genetic Algorithm, was adapted to evaluate >64,000 potential polynomial equations as predictors.

Results: Of the 4607 eligible women, 4321 (94%) had a pregnancy without major complications and delivered a live singleton without congenital malformations. After other exclusions (missing measurements in GA window and outliers), the final sample comprised 4229 women. Two skeletal measures, HC and FL, produced the best GA prediction, given by the equation

$\log_e(GA) = 0.03243 \times [\log_e(HC)]^2 + 0.001644 \times FL \times \log_e HC + 3.813$

When FL is not available, the best equation based on HC alone is

$\log_e(GA) = 0.05970 \times [\log_e(HC)]^2 + 0.000000006409 \times HC^3 + 3.3258$

The estimated uncertainty of GA prediction (95% interval) was 6 to 7 days at 14 weeks' gestation, 12 to 14 days at 26 weeks' gestation, and over 14 days in the third trimester. The addition of FL to HC leads to improved prediction intervals over just using HC, but no further improvement in prediction is afforded by adding AC, BPD or OFD. Equations that included other measurements (BPD, OFD and AC) did not perform better.

Conclusions: Among women initiating antenatal care late in pregnancy a single set of ultrasound measurements combining HC and FL in the second trimester can be used to estimate GA with reasonable accuracy. We recommend this tool for underserved populations but considerable efforts should be implemented to improve early initiation of antenatal care worldwide.

INTRODUCTION

Reliable estimation of gestational age (GA) is essential as it allows appropriate scheduling of a woman's antenatal care, informs obstetric management decisions and facilitates the correct interpretation of fetal growth assessment [1]. Abnormal fetal growth patterns such as growth restriction or macrosomia may be missed or incorrectly diagnosed if GA is unknown or incorrect. Reliable GA estimation is also important at a population level to calculate rates of preterm birth and small for gestational age (SGA) at birth. The lack of accurate GA estimation, particularly in geographical regions at greatest risk of these conditions, means that preterm birth and SGA rates are mere approximations in many parts of the world [2, 3].

Traditionally, GA is estimated using the first day of the last menstrual period (LMP), which assumes that ovulation occurs on day 14 of the menstrual cycle. Irregular menses, unknown or uncertain dates, oral contraceptive use or recent pregnancy or breastfeeding may all influence the accuracy of this method - issues that occur in a large proportion of women [4-6]. In such cases, early (<14 weeks' gestation) ultrasound measurement of fetal crown rump length (CRL) is recommended [7,8]. First trimester GA assessment is more accurate than dating in late pregnancy because, with advancing gestation, fetal ultrasound measurements have a larger absolute error [9] and growth disturbances become more noticeable, resulting in potential underestimation of GA for an abnormally small fetus and overestimation for a macrosomic fetus.

Unfortunately, in many settings where high-risk pregnancies are prevalent, women attend their first antenatal care visit late in pregnancy or even at the time of delivery. This makes it difficult to manage complications, evaluate fetal growth and implement evidence-based interventions, such as the administration of corticosteroids for fetal lung maturation in cases of threatened preterm labour.

The present analysis of the Fetal Growth Longitudinal Study (FGLS), one of the main components of the INTERGROWTH-21st Project, aims to complement our previous work of early GA estimation by ultrasound measurement of CRL [10]. We have explored a set of equations to estimate GA using fetal biometry acquired during a single ultrasound scan performed between 14 and 34 weeks' gestation.

METHODS

INTERGROWTH-21st is a multicentre, multiethnic, population-based project, conducted between 2009 and 2014 in eight countries [11]. Its primary aim was to study growth, health, nutrition and neurodevelopment from <14 weeks' gestation to 2 years of age, using the same conceptual framework as the WHO Multicentre Growth Reference Study (MGRS) [12,13].

Eight urban areas located at low altitude (≤1,600m) were chosen as study sites, within which we selected all institutions providing pregnancy and intrapartum care where >80% of deliveries occurred. Women receiving antenatal care had to plan to deliver in these institutions or in a similar hospital located in the same geographical area and there had to be an absence or low levels of major, known, non-microbiological contamination such as pollution, domestic smoke, radiation or any other toxic substances [14].

Women from these populations, with a singleton pregnancy that was conceived naturally, and who met the individual inclusion criteria, were recruited prospectively and consecutively into FGLS. The study methods have been described in detail elsewhere [11,15].

The true gestational age (GA_{true}) was defined by the woman's LMP determined at the first visit at $<14^{+0}$ weeks' gestation, provided that: (1) the date was certain; (2) she had a regular 24–32 day menstrual cycle; (3) she had not been using hormonal contraception or breastfeeding in the preceding 2 months, and (4) it was in agreement (within 7 days) with a measurement of fetal CRL at 9^{+0} to 13^{+6} weeks' gestation [15].

A single ultrasound machine (Philips HD-9; Philips Ultrasound, Bothell, WA, USA) with curvilinear abdominal transducers (C5-2, C6-3, V7-3) was used for all fetal measurements $\geq 14^{+0}$ weeks' gestation. To reduce expected value bias, the ultrasound machines were specially adapted so that the measurements were not visible on the screen. However, as women presented for their first visit at different clinics within the geographical area, for those ultrasound scans done $<14^{+0}$ weeks' gestation (CRL measures only), it was considered acceptable to use other, locally available, machines provided they were evaluated and approved by the study team. All ultrasonographers (n=39) at the eight study sites underwent rigorous training and standardization. In accordance with the protocol, CRL and fetal biometry measures were assessed for quality; the former were also reviewed blindly by our collaborators at the Société Française pour l'Amélioration des Pratiques Echographiques (SFAPE) [16,17].

Women were invited for follow-up ultrasound scans every 5 weeks (within one week either side) after the initial dating scan, so that the possible ranges after the dating scan were 14-18, 19-23, 24-28, 29-33, 34-38 and 39-42 weeks' gestation. At each visit, fetal head circumference (HC), biparietal diameter (BPD), occipitofrontal diameter (OFD), abdominal circumference (AC) and femur length (FL) were measured three times from three separately obtained ultrasound images of each structure [18].

Head measurements were taken in an axial view at the level of the thalami, with an angle of insonation as close as possible to 90°. The head had to be oval in shape, symmetrical, centrally positioned and filling at least 30% of the monitor. The midline echo (representing the falx cerebri) had to be broken anteriorly, at one-third of its length, by the cavum septi pellucidi. The thalami had to be located symmetrically on either side of the midline. Callipers were then placed on the outer border of the parietal bones ('outer to outer') at the widest or longest part of the skull for the BPD and OFD, respectively; the HC was measured using the ellipse facility on the outer border of the skull.

The measurements of the fetal abdomen were taken in a cross-sectional view (as close as possible to a circle), with the umbilical vein in the anterior third of the abdomen (at the level of the portal sinus), and the stomach bubble visible. The operator was instructed to avoid applying too much pressure with the transducer as this can distort the circular shape of the fetal abdomen. The abdomen had to fill at least 30% of the monitor screen; the spine had preferably to be positioned at either 3 or 9 o'clock to avoid internal shadowing; the kidneys and bladder had not to be visible. For the measurements, the contour of the ellipse was placed on the outer border of the abdomen.

Finally, the FL was measured using a longitudinal view of the fetal thigh closest to the probe and with the femur as close as possible to the horizontal plane. The angle of insonation of the ultrasound beam was approximately 90° with the full length of the bone visualised, unobscured by shadowing from adjacent bony parts, and the femur had to fill at least 30% of the monitor screen. The intersection of the callipers was placed on the outer borders of the edges of the femoral diaphysis (outer to outer) ensuring clear femoral edges.

Detailed measurement protocols, standardization procedures and quality control methods employed across all sites are described in detail elsewhere [15,19-21].

Statistical analysis

For each woman included in the study, a single ultrasound scan between 14+0 and 40+0 weeks' gestation was randomly selected using the 'sample' function in Stata (version 13). At each scan, the routinely measured fetal biometric variables were recorded. To overcome the problem of data truncation at the lower end of gestation (<14+0 weeks), we followed the same approach previously described and applied to CRL data [22]. Using the international fetal growth equations for head circumference, abdominal circumference, femur length, biparietal diameter and occipito-frontal diameter [18], we simulated 20 observations for each day between 12+0 and 13+6 weeks' gestation (n=280), which is approximately the same number of observations for each day of GA in the un-truncated data set. After simulation, we restricted the data based on head circumference by excluding values below 85mm or above 330mm and visually inspecting a plot of the data to assess that the truncation problem had been overcome. Using the augmented data set, fractional polynomial regression analyses were employed using the Xrigls function in Stata to model the mean and standard deviation (SD) of GA for each biometric variable [22].

In order to establish the relationship of fetal biometric variables and gestational age we used an automated machine learning, 'Genetic Algorithm' (see Appendix). This method was chosen because a more traditional fractional polynomial approach, which is well suited to modelling a single variable, has limited scope when used with multiple biometric variables that are highly correlated. By virtue of the automated approach, the Genetic Algorithm is able to evaluate large numbers (in this case over 64,000) of potential combinations of biometric variables that are used to build polynomial equations as predictors of GA; this would not be feasible using conventional approaches. By specifying a mathematical definition of optimal performance, based upon minimization of the mean prediction error (root mean squared error, RMSE), the first stage of model development was entirely automated with the capacity to assemble, evaluate and modify equations. We were, therefore, able to use the data themselves to generate preliminary models in an entirely objective manner.

Briefly, a large number of preliminary candidate equations were developed using combinations of all candidate biometric variables (including powers (0.5, 1, 2 and 3), their logs and their products). Each of the candidate equations was used to obtain for each fetus a predicted gestational age ($GA_{predicted}$) as an estimate of their true gestational age (GA_{true}). After preliminary analysis it was clear that the predicted values $GA_{predicted}$ were not normally

distributed; this was addressed by predicting the natural logarithm of $GA_{predicted}$ ($\log_e GA_{predicted}$). The equations were then ranked to assess which had the lowest uncertainty based on the 95% prediction interval.

We used a 4-step approach to determine our final equation (see Appendix for more detailed explanation):

1. **Equation discovery using the Genetic Algorithm:** The automated Genetic Algorithm was used to determine the equation providing the best prediction of GA using combinations of fetal biometric variables. Briefly, the model initiated itself by assigning polynomial equations linking fetal biometry within the dataset and GA_{true}. Model terms, coefficients and powers were randomly selected within specified limits**.** Once defined, the individual equations were each used to predict GA_{true} using the observed fetal biometry data. The performance of individual equations was measured by calculating the RMSE between the true and predicted GA at each iteration of the genetic algorithm. For each combination of biometric variables, the equation with the lowest RMSE was selected automatically and modified by methods that mimic the genetic principles of mutation and cross-over. Thus, a second generation of equations was developed with the 'positive predictive qualities' of the first generation preserved in their structure. Furthermore, random variation was introduced as 'mutations' into the best performing equations in order to assess whether such mutations conferred an advantage in prediction, evaluated using the RMSE. By repeating the process over many iterations, the structure of equations was continuously refined until there was convergence upon the equation, or series of equations, that most accurately predicted GA_{true}. All data processing at this stage was performed using the `GAPolyfitn' function in MATLAB version R2014b.

2. **Goodness of fit.**

   Visual inspection of scatter plots was used to compare GA_{true} with $GA_{predicted}$ for each candidate equation obtained from the genetic algorithm. Quantile-Quantile (QQ) plots were used to compare the distributions. Well-fitting models were identified by a QQ plot with minimal deviation from the line of equality.

   For each equation, absolute residuals between $GA_{true} - GA_{predicted}$ were regressed on GA using fractional polynomial methods (powers +/- 0.5, 1, 2 and 3) to provide an equation that approximated the SD, and multiplied by a constant to estimate the 2.5th

and 97.5th centiles using the Xrigls function in Stata (Altman DG. Construction of age-related reference centiles using absolute residuals. Stat Med 1993;12:917-924). The goodness of fit of these estimated SDs was assessed by calculating the proportion of predicted GAs that were outside the 95% prediction intervals (±1.96SD), which should be 5%.

3. Evaluation of model complexity

To facilitate a suitable balance between parsimony and model performance, estimates of Akaike's Information Criterion (AIC) [23] were calculated and compared. The AIC combines an estimate of the goodness of fit of a model with a penalty for increasing model complexity. In addition, candidate equations with similar indices were compared in terms of number of terms and complexity, defined as the sum of the powers of each variable). Where two equations demonstrated similar performance, the equation with a less complex structure was preferred.

4. Post-production model refinement

After examining the model complexity, it appeared that most of the contribution to the prediction of GA was based on HC. Therefore, simplified models were constructed, restricted to biometry of the fetal head (HC, OFD and BPD).

The INTERGROWTH-21st Project was approved by the Oxfordshire Research Ethics Committee "C" (ref: 08/H0606/139), the research ethics committees of the individual participating institutions, as well as the corresponding regional health authorities where the project was implemented.

Results

Of the 4,607 women recruited into FGLS, 4,321 delivered live singletons without congenital malformations (Figure 1). Exclusions were women with missing fetal measurements (n=84), and outliers defined as fetal measurements >5 SD above/below the mean (n=7). We further restricted the actual data at the top end by excluding head circumference values above 330mm (n=1) resulting in a total of 4,229 women who contributed a single, randomly selected ultrasound scan between 14+0 and 40+0 weeks' gestation (Figure 1). Of the 280 observations simulated and added to actual data, we similarly excluded head circumference values below 85mm (n=148) to obtain the final analysis sample of 4,361 observations. The baseline characteristics and perinatal events of the study population (n=4,229 excluding simulated observations) are shown in Table 1.

The equations that best estimated GA_{true} based on lowest RMSE, best fit and optimal AIC are shown in Table 2. Equations selected were based on HC alone, and a combination of HC and FL; despite including multiple measures (HC, BPD, OFD, AC, FL) in the models, only HC and FL were retained after the selection process.

Overall, based on a model using HC only, the uncertainty of estimated GA gradually with advancing GA, from 6 to 7 days in either direction at 14^{+0} weeks to 14 to 18 days at 32^{+0} weeks' gestation (Table 3). Inclusion of FL led to an improvement in prediction throughout gestation of about 1-3 days. Inclusion of the other parameters led to no further improvement, and so none were included in the equations resulting from the Genetic Algorithm search.

The plots of $GA_{predicted}$ versus GA_{true} between 14^{+0} and 34^{+0} weeks' gestation demonstrated good model fitting (Figures 2 and 3 for the scatterplots and Figures S4 and S5 for the QQ plots).

Apart from estimating the most likely GA (by using $GA_{predicted}$ for a set of measurements), we also present the lower and upper bounds of the estimation of GA (Table 4). The lower bound can be used in clinical management for women who present in late pregnancy: it is an estimate of the likely "least GA", e.g. 97.5% likely to be at least X weeks.

Discussion

Main findings

We have shown that a single set of basic ultrasound measurements of HC and FL in the second trimester can be used to estimate GA with reasonable accuracy. The estimation is best at lower GAs where the 95% prediction interval is within 6 days, but is just over 12 days at 26^{+0} weeks' gestation. The addition of FL to HC leads to considerable improvement over just using HC, but no further improvement in prediction is afforded by adding AC, BPD or OFD.

Strengths and weaknesses

We have produced equations for GA assessment that are more precise than those currently used in routine clinical practice (Table 5). This may be due to the prospective nature of the study; a large sample size; accurately dated pregnancies; a clearly defined measurement protocol; quality control measures, and a statistical approach that searched for the optimal combination of factors iteratively, rather than relying upon a user-controlled search. The multicentre, international setting of the study with measurements taken by a large group of ultrasonographers provides external validity.

There is an intrinsic limitation when estimating GA by fetal anthropometric based equations: namely that what is measured is fetal size not GA. Fetal size may vary for reasons other than differences in GA especially as factors conditioning abnormal fetal growth are more prevalent in the populations where the equation is most likely to be used. In other words, it is important to take into account the impact of pathology (fetal growth restriction and overgrowth) on GA estimation. This is true for any equation estimating GA – the accuracy at an individual level will depend on the 'normality' of the fetal size and, at the population level, on the prevalence of abnormal growth patterns. Thus, efforts should focus on modifying health systems and referral pathways to prevent late presentation in pregnancy, rather than simply achieving technological advances in fetal size-based dating.

Interpretation

Ultrasound assessment of GA is performed assuming that fetal size can be used as a proxy for GA. This assumption depends on: a) the GA at which biometry is performed (at earlier GAs growth is more uniform and there is less measurable growth impairment); b) the choice of biometric variable, and c) the accuracy of the measurement, which is affected by technical aspects of imaging and operator skill.

The most accurate way to estimate GA is by measuring fetal CRL between 8^{+0} and 14^{+0} weeks' gestation, which is associated with a 95% prediction interval of 2.7 days [24,10]. This method is the basis of recommended pregnancy dating policies throughout much of the developed world [25]. Beyond 14^{+0} weeks' gestation, fetal flexion limits the accuracy of CRL measurements for dating purposes and GA estimates are based on measurement of the HC, BPD, AC, FL or a combination of these [8, 26].

Our results demonstrate the relative inaccuracy of late GA assessment, which is due to the increasing biological variability in fetal size as well as the increasing absolute error of fetal measurements with advancing GA [9]. Therefore, all information (clinical and imaging) should be considered when dating pregnancies and providing obstetric care – particularly after late pregnancy dating. Thus, we recommend that the following principles should be applied in clinical practice:

- Assessment of fetal age should be based on the earliest available ultrasound measurement after 8^{+0} weeks' gestation, provided the measurements are technically adequate. CRL should be used before 14^{+0} weeks' gestation and equation 2 (HC and FL) after 14^{+0} weeks but before 26^{+0} weeks' gestation. In settings where HC only is available then equation 1 can be used.
- If menstrual dates are reliable and within the prediction limits of the fetal measurement, ultrasound should merely confirm the GA assessed by LMP.
- When menstrual dates are reliable and fall outside the prediction interval of ultrasound assessment there are two interpretations: the menstrual dates are incorrect and GA should be based on ultrasound measurement; or the GA is correct as assessed by LMP and the fetus is an abnormal size for that GA (or both). Clinical features of growth restriction, e.g. reduced amniotic fluid or abnormal uterine or umbilical artery blood flow, should be taken into consideration, as should factors that may lead to overgrowth, e.g. maternal diabetes. An interval ultrasound scan should then be carried out to confirm GA.
- When menstrual dates are unknown, GA estimation should be based on ultrasound, which has reasonable accuracy until 26^{+0} weeks' gestation; and a further ultrasound scan should be carried out.

Finally, when GA is estimated in the third trimester, the possible error is large and must be taken into account to ensure safe obstetric practice. The use of the concept of a "minimum" GA, by using the lower limit of the prediction interval from the equation can be useful in this instance (Table 4). For example, if a woman presents in threatened preterm labour and the ultrasound examination suggests a median GA of 34^{+6} based on HC and FL, it should be appreciated that the GA could be as low as 32^{+6} weeks. i.e. "it is most likely that the GA is 34^{+6} weeks, but we are 95% certain that the GA is *at least* 32^{+6} weeks. However, if the fetus is growth restricted, the GA could be as much as 36^{+5} weeks." This analysis is very relevant to clinical decision making: for example, when administering prophylactic corticosteroids or transferring a neonate to a higher level of care. In contrast, labour induction may be considered at 40^{+0} weeks' gestation based on late assessment, as the GA could be more advanced. Such a clinically cautious approach is particularly important as it is known that unreliable reporting of LMP and late antenatal care are both associated with adverse pregnancy outcomes [5, 27].

Conclusion

We have shown that a single set of ultrasound measurements in the second trimester can be used to estimate GA with relative accuracy. We recommend these tools for the management of women who present late in pregnancy. However, we strongly encourage as a priority the promotion of early antenatal care in regions and sub-populations that are not yet benefitting from this practice.

REFERENCES

1. Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A, Papageorghiou AT. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. BJOG. 2012 Nov;119(12):1425-39
2. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A, Narwal R, Adler A, Vera Garcia C, Rohde S, Say L, Lawn JE. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. The Lancet. 2012; 379(9832):2162–2172
3. de Onis , Blössner M, Villar J. Levels and patterns of intrauterine growth retardation in developing countries. Eur J Clin Nutr. 1998;52 Suppl 1:S5-15.
4. Savitz DA, Terry JW, Dole N, Thorp JM, 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.
5. Thorsell M, Kaijser M, Almström H, Andolf E. Expected day of delivery from ultrasound dating versus last menstrual period – obstetric outcome when date mismatch. BJOG 2008;115:585–9.
6. Nguyen TH, Larsen T, Engholm G, Moller H. Evaluation of ultrasoundestimated date of delivery in 17,450 spontaneous singleton births: do we need to modify Naegele's rule? Ultrasound Obstet Gynecol. 1999;14:23–8.
7. Hughes R, Aitken E, Anderson J, Barry C, Benton M, Elliot J; National Institute for Health and Clinical Excellence. Antenatal care. Routine care for the healthy pregnant woman. NICE clinical guideline 62. London: RCOG Press; 2008.
8. Butt K, Lim K; Society of Obstetricians and Gynaecologists of Canada Clinical Practice Guideline. Determination of Gestational Age by Ultrasound. J Obstet Gynaecol Can 2014;36(2):171–181
9. Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L, Altman DG, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Intra- and interobserver variability in fetal ultrasound measurements. Ultrasound Obstet Gynecol. 2012;39(3):266-73.

10. Papageorghiou AT, Kennedy SH, Salomon LJ, Ohuma EO, Cheikh Ismail L, Barros FC, Lambert A, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Altman DG, Purwar M, Noble JA, Pang R, Victora CG, Bhutta ZA, Villar J; for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester of pregnancy. Ultrasound Obstet Gynecol. 2014 Jul 8

11. Villar J, Altman DG, Purwar M, Noble JA, Knight HE, Ruyan P, Cheikh Ismail L, Barros FC, Lambert A, Papageorghiou AT, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Bhutta ZA, Kennedy SH; International Fetal and Newborn Growth Consortium for the 21st Century. The objectives, design and implementation of the INTERGROWTH-21st Project. BJOG. 2013;120 Suppl 2:9-26, v.

12. Garza C, de Onis M. Rationale for developing a new international growth reference. Food Nutr Bull 2004; 25 (suppl): S5–14

13. de Onis M, Garza C, Onyango AW, Martorell R. WHO child growth standards. Acta Paediatr Suppl 2006; 450: 1–101.

14. Eskenazi B, Bradman A, Finkton D, Purwar M, Noble JA, Pang R, Burnham O, Cheikh Ismail L, Farhi F, Barros FC, Lambert A, Papageorghiou AT, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Altman DG, Ohuma EO, Kennedy SH, Bhutta ZA, Villar J; International Fetal and Newborn Growth Consortium for the 21st Century. A rapid questionnaire assessment of environmental exposures to pregnant women in the INTERGROWTH-21st Project. BJOG 2013; 120 (suppl 2): 129-38

15. Papageorghiou AT, Sarris I, Ioannou C, Todros T, Carvalho M, Pilu G, Salomon LJ; International Fetal and Newborn Growth Consortium for the 21st Century. Ultrasound methodology used to construct the fetal growth standards in the INTERGROWTH-21st Project. BJOG 2013; 120 (suppl 2): 27–32.

16. Ioannou C, Sarris I, Hoch L, Salomon LJ, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century. Standardisation of crown-rump length measurement. BJOG 2013; 120 (suppl 2): 38–41.

17. Wanyonyi S, Napolitano R, Ohuma E, Salomon L, Papageorghiou A. Image-scoring system for crown-rump length measurements. Ultrasound Obstet Gynecol. 2014 Mar 27. doi: 10.1002/uog.13376

18. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, Jaffer YA, Bertino E, Gravett MG, Purwar M, Noble JA, Pang R, Victora CG, Barros FC, Carvalho M, Salomon LJ, Bhutta ZA, Kennedy SH, Villar J; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet. 2014 Sep 6;384(9946):869-79.

19. Sarris I, Ioannou C, Ohuma EO, Altman DG, Hoch L, Cosgrove C, Fathima S, Salomon LJ, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. BJOG 2013; 120 (suppl 2): 33–37.

20. Sarris I, Ioannou C, Dighe M, Mitidieri A, Oberto M, Qingqing W, Shah J, Sohoni S, Al Zidjali W, Hoch L, Altman DG, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. Ultrasound Obstet Gynecol 2011; 38: 681–87.

21. www.intergrowth21.org.uk

22. Ohuma EO, Papageorghiou AT, Villar J, Altman DG. Estimation of gestational age in early pregnancy from crown-rump length when gestational age range is truncated: the case study of the INTERGROWTH-21st Project. BMC Med Res Methodol. 2013 Dec 7;13:151. doi: 10.1186/1471-2288-13-151.

23. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19:716e23

24. Robinson HP. Sonar measurement of fetal crown-rump length as means of assessing maturity in first trimester of pregnancy. British Medical Journal. 1973; 4(5883):28–31

25. Napolitano R, Dhami J, Ohuma EO, Ioannou C, Conde-Agudelo A, Kennedy SH, Villar J, Papageorghiou AT. Pregnancy dating by fetal crown-rump length: a systematic review of charts. BJOG. 2014 Apr;121(5):556-65. doi: 10.1111/1471-0528.12478.

26. Salomon LJ, Alfirevic Z, Bilardo CM, Chalouhi GE, Ghi T, Kagan KO, Lau TK, Papageorghiou AT, Raine-Fenning NJ, Stirnemann J, Suresh S, Tabor A, Timor-Tritsch IE, Toi A, Yeo G. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol. 2013 Jan;41(1):102-13. doi: 10.1002/uog.12342.

27. Nguyen TH, Larsen T, Engholm G, Møller H. Increased adverse pregnancy outcomes with unreliable last menstruation. Obstet Gynecol. 2000 Jun;95(6 Pt 1):867-73.

28. Hadlock FP, Deter RL, Harrist RB, Park SK. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. Radiology. 1984 Aug;152(2):497-501.

29. Altman DG, Chitty LS. New charts for ultrasound dating of pregnancy. Ultrasound Obstet Gynecol 1997;10:174–191

Financial Disclosure Statement: The authors have indicated that they have no financial relationships relevant to this article to disclose.

Conflict of Interest Statement: The authors have indicated that they have no conflict of interest to disclose.

Acknowledgements

This project was supported by a generous grant from the Bill & Melinda Gates Foundation to the University of Oxford, for which we are very grateful. We would also like to thank the Health Authorities in Pelotas, Brazil; Beijing, China; Nagpur, India; Turin, Italy; Nairobi, Kenya; Muscat, Oman; Oxford, UK and Seattle, USA, who facilitated the project by allowing participation of these study sites as collaborating centres. We are extremely grateful to Philips Medical Systems who provided the ultrasound equipment and technical assistance throughout the project. We also thank MedSciNet U.K. Ltd for setting up the INTERGROWTH-21st web-site and for the development, maintenance and support of the on-line data management system.

We thank the parents and infants who participated in the studies and the more than 200 members of the research teams who made the implementation of this project possible. The participating hospitals included: Brazil, Pelotas (Hospital Miguel Piltcher, Hospital São Francisco de Paula, Santa Casa de Misericórdia de Pelotas, and Hospital Escola da Universidade Federal de Pelotas); China, Beijing (Beijing Obstetrics & Gynecology Hospital, Shunyi Maternal & Child Health Centre, and Shunyi General Hospital); India, Nagpur (Ketkar Hospital, Avanti Institute of Cardiology Private Limited, Avantika Hospital, Gurukrupa Maternity Hospital, Mulik Hospital & Research Centre, Nandlok Hospital, Om Women's Hospital, Renuka Hospital & Maternity Home, Saboo Hospital, Brajmonhan Taori Memorial Hospital, and Somani Nursing Home); Kenya, Nairobi (Aga Khan University Hospital, MP Shah Hospital and Avenue Hospital); Italy, Turin (Ospedale Infantile Regina Margherita Sant' Anna and Azienda Ospedaliera Ordine Mauriziano); Oman, Muscat (Khoula Hospital, Royal Hospital, Wattayah Obstetrics & Gynaecology Poly Clinic, Wattayah Health Centre, Ruwi Health Centre, Al-Ghoubra Health Centre and Al-Khuwair Health Centre); UK, Oxford (John Radcliffe Hospital) and USA, Seattle (University of Washington Hospital, Swedish Hospital, and Providence Everett Hospital).

Full acknowledgement of all those who contributed to the development of the INTERGROWTH-21st Project protocol appears at www.intergrowth21.org.uk

Appendix : Members of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) and its Committees

Scientific Advisory Committee

M Katz (Chair from January 2011), MK Bhan, C Garza, S Zaidi, A Langer, PM Rothwell (from February 2011), Sir D Weatherall (Chair until December 2010).

Steering Committee

ZA Bhutta (Chair), J Villar (Principal Investigator), S Kennedy (Project Director), DG Altman, FC Barros, E Bertino, F Burton, M Carvalho, L Cheikh Ismail, WC Chumlea, MG Gravett, YA Jaffer, A Lambert, P Lumbiganon, JA Noble, RY Pang, AT Papageorghiou, M Purwar, J Rivera, C Victora.

Executive Committee

J Villar (Chair), DG Altman, ZA Bhutta, L Cheikh Ismail, S Kennedy, A Lambert, JA Noble, AT Papageorghiou.

Project Coordinating Unit

J Villar (Head), S Kennedy, L Cheikh Ismail, A Lambert, AT Papageorghiou, M Shorten, L Hoch (until May 2011), HE Knight (until August 2011), EO Ohuma (from September 2010), C Cosgrove (from July 2011), I Blakey (from March 2011).

Data Analysis Group

DG Altman (Head), EO Ohuma, J Villar.

Data Management Group

DG Altman (Head), F Roseman, N Kunnawar, SH Gu, JH Wang, MH Wu, M Domingues, P Gilli, L Juodvirsiene, L Hoch (until May 2011), N Musee (until June 2011), H Al-Jabri (until October 2010), S Waller (until June 2011), C Cosgrove (from July 2011), D Muninzwa (from October 2011), EO Ohuma (from September 2010), D Yellappan (from November 2010), A Carter (from July 2011), D Reade (from June 2012), R Miller (from June 2012).

Ultrasound Group

AT Papageorghiou (Head), LJ Salomon (Senior external advisor), A Leston, A Mitidieri, F Al-Aamri, W Paulsene, J Sande, WKS Al-Zadjali, C Batiuk, S Bornemeier, M Carvalho, M Dighe, P Gaglioti, N Jacinta, S Jaiswal, JA Noble, K Oas, M Oberto, E Olearo, MG Owende, J Shah, S Sohoni, T Todros, M Venkataraman, S Vinayak, L Wang, D Wilson, QQ Wu, S Zaidi, Y Zhang, P Chamberlain (until September 2012), D Danelon (until July 2010), I Sarris (until June 2010), J Dhami (until July 2011), C Ioannou (until February 2012), CL Knight (from October 2010), R Napolitano (from July 2011), S Wanyonyi (from May 2012), C Pace (from January 2011), V Mkrtychyan (from June 2012).

Anthropometry Group

L Cheikh Ismail (Head), WC Chumlea (Senior external advisor), F Al-Habsi, ZA Bhutta, A Carter, M Alija, JM Jimenez-Bustos, J Kizidio, F Puglia, N Kunnawar, H Liu, S Lloyd, D Mota, R Ochieng, C Rossi, M Sanchez Luna, YJ Shen, HE Knight (until August 2011), DA Rocco (from June 2012), IO Frederick (from June 2012).

Neonatal Group

ZA Bhutta (Head), E Albernaz, M Batra, BA Bhat, E Bertino, P Di Nicola, F Giuliani, I Rovelli, K McCormick, R Ochieng, RY Pang, V Paul, V Rajan, A Wilkinson, A Varalda (from September 2012).

Environmental Health Group

B Eskenazi (Head), LA Corra, H Dolk, J Golding, A Matijasevich, T de Wet, JJ Zhang, A Bradman, D Finkton, O Burnham, F Farhi.

Participating countries and local investigators

Brazil: FC Barros (Principal Investigator), M Domingues, S Fonseca, A Leston, A Mitidieri, D Mota, IK Sclowitz, MF da Silveira.

China: RY Pang (Principal Investigator), YP He, Y Pan, YJ Shen, MH Wu, QQ Wu, JH Wang, Y Yuan, Y Zhang.

India: M Purwar (Principal Investigator), A Choudhary, S Choudhary, S Deshmukh, D Dongaonkar, M Ketkar, V Khedikar, N Kunnawar, C Mahorkar, I Mulik, K Saboo, C Shembekar, A Singh, V Taori, K Tayade, A Somani.

Italy: E Bertino (Principal Investigator), P Di Nicola, M Frigerio, G Gilli, P Gilli, M Giolito, F Giuliani, M Oberto, L Occhi, C Rossi, I Rovelli, F Signorile, T Todros.

Kenya: W Stones and M Carvalho (Co- Principal Investigators), J Kizidio, R Ochieng, J Shah, , S Vinayak, N Musee (until June 2011), C Kisiang'ani (until July 2011), D Muninzwa (from August 2011).

Oman: YA Jaffer (Principal Investigator), J Al-Abri, J Al-Abduwani, FM Al-Habsi, H Al-Lawatiya, B Al-Rashidiya, WKS Al-Zadjali, FR Juangco, M Venkataraman, H Al-Jabri (until October 2010), D Yellappan (from November 2010).

UK: S Kennedy (Principal Investigator), L Cheikh Ismail, AT Papageorghiou, F Roseman, A Lambert, EO Ohuma, S Lloyd, R Napolitano (from July 2011), C Ioannou (until February 2012), I Sarris (until June 2010).

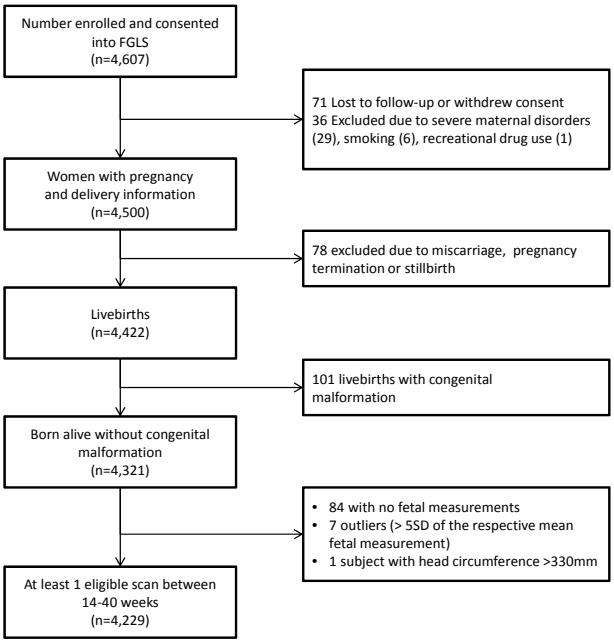
USA: MG Gravett (Principal Investigator), C Batiuk, M Batra, S Bornemeier, M Dighe, K Oas, W Paulsene, D Wilson, IO Frederick, HF Andersen, SE Abbott, AA Carter, H Algren, DA Rocco, TK Sorensen, D Enquobahrie, S Waller (until June 2011).

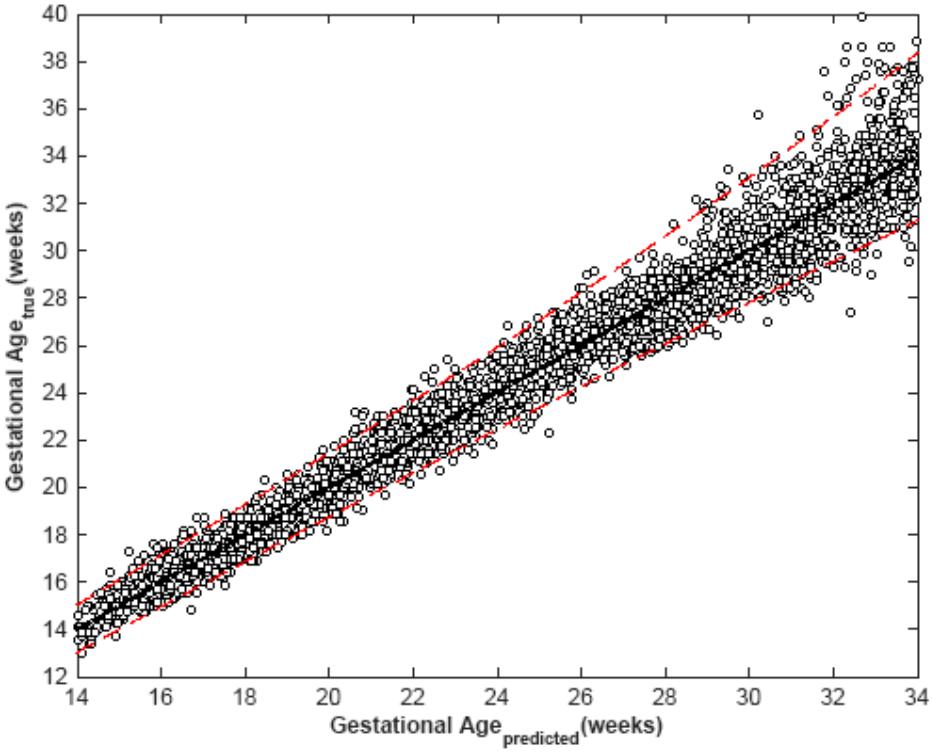
Figure 1: Flow chart of recruitment through the study.

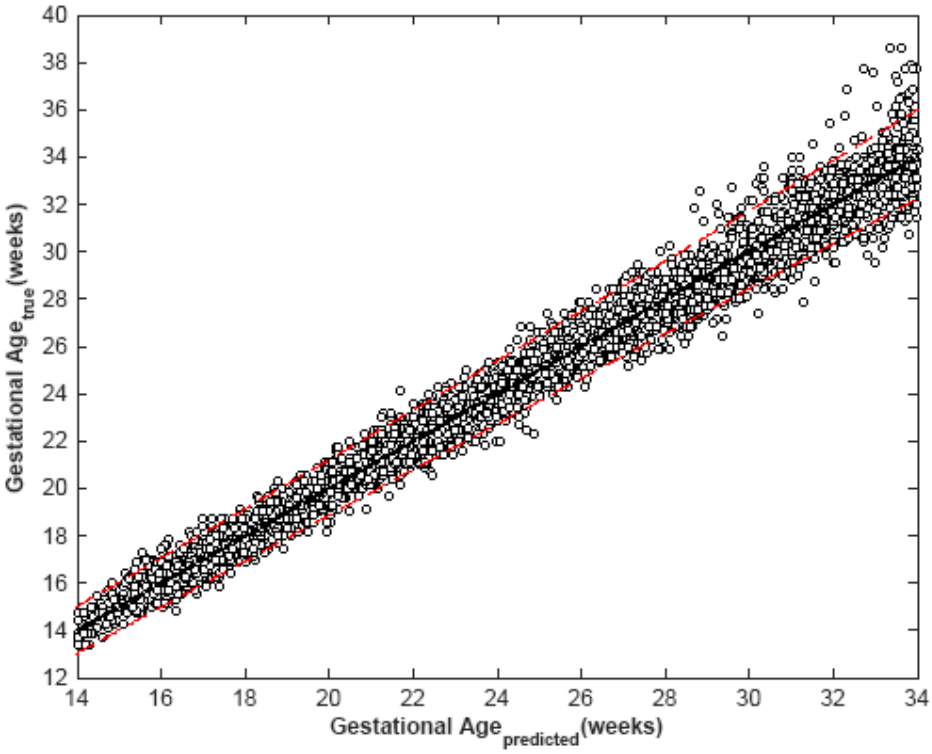
Figure 2: Scatterplot showing the estimation of gestational age based on fetal head circumference at 14 to 34 weeks of true gestational age. The black line is the line of equality; the red lines are ±2SD.

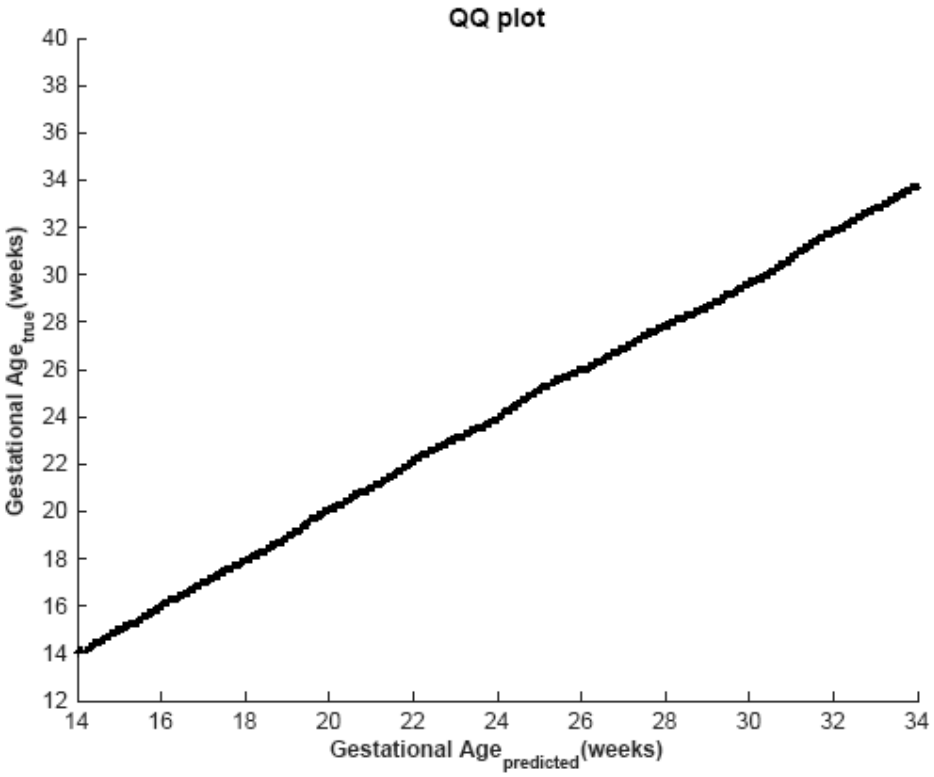
Figure 3: Scatterplot showing estimation of gestational age based on fetal head circumference and femur length at 14 to 34 weeks of true gestational age. The black line is the line of equality; the red lines are ±2SD.

Supplementary figures S4-S5: Q-Q plots to assess the goodness of fit of the models for head circumference (S4); and head circumference and femur length (S5).









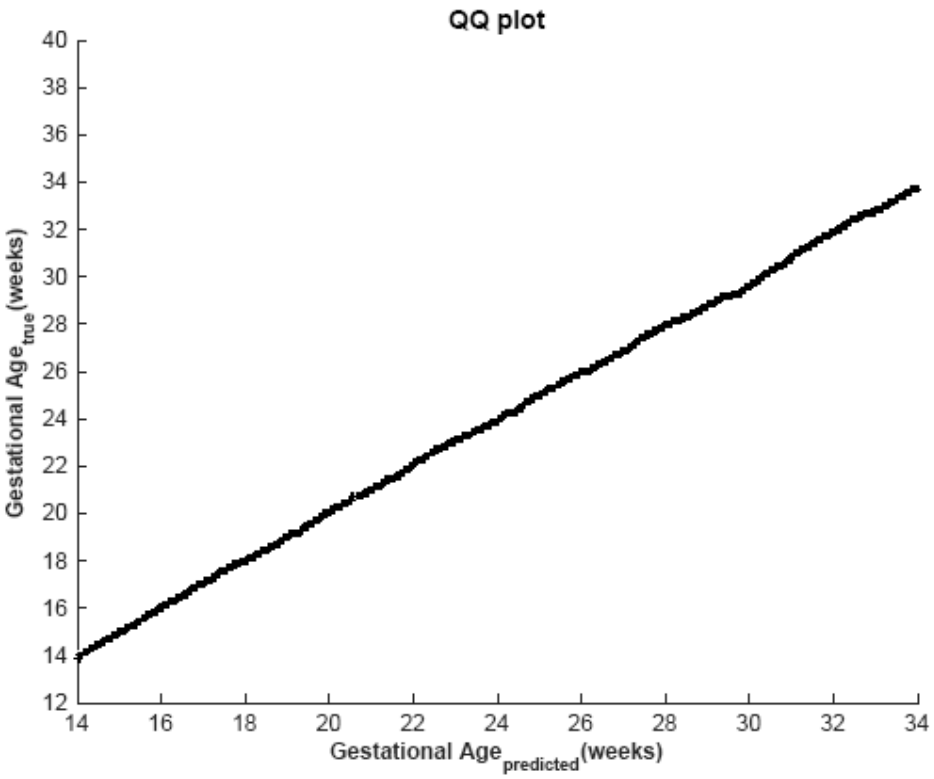


Table 1: Baseline characteristics and perinatal events of the study population of 4229 women

Characteristic	**Mean (SD) or n (%)**
Maternal age, years	27.8 (3.8)
Maternal height, cm	162.2 (5.8)
Maternal weight, kg	61.5 (9.2)
Body mass index, kg/m^2	23.2 (3.0)
Gestational age at first visit, weeks	11.3 (1.4)
Nulliparous	2815 (68.6%)
Pre-eclampsia	31 (<1%)
Preterm birth (<37 weeks)	189 (4.5%)
Term LBW less than 2500g (≥37 weeks)	127 (3%)
Male sex	2101 (49.7%)
Birth weight (≥37 weeks), kg	3.3 (0.4)

Maternal baseline characteristics were measured at less than 14 weeks' gestation. LBW = low birth weight.

Table 2: Selected equations for late gestational age estimation

ID	Variables (mm)	Equation to estimate $\log_e$GA (days)
1	HC	$\log_e(GA) = 0.05970 \times [\log_e(HC)]^2 + 0.000000006409 \times HC^3 + 3.3258$
2	HC, FL	$\log_e(GA) = 0.03243 \times [\log_e(HC)]^2 + 0.001644 \times FL \times \log_e(HC) + 3.813$
ID	**Variables (mm)**	**Equation to estimate the standard deviation of $GA_{predicted}$ (days)**
1	HC	$SD = 0.6492 \times (GA \times 0.01)^3 + 2.991$
2	HC, FL	$SD = 0.04009 \times GA - 1.149$

GA: exact estimated gestational age (days)
HC: Head Circumference (mm)
FL: Femur Length (mm)
$\log_e$ = Logarithm to base e (natural logarithm)
SD = standard deviation

For example, to calculate gestational age using equation 1:
If HC is 250mm
median GA = exp $(0.05970 * [\log_e(250)]^2 + 0.000000006409 * 250^3 + 3.3258$
= exp (5.245986506)
= 189.8 days (equivalent to 27.1 weeks)

To calculate gestational age using equation 2,
If HC is 250mm and FL is 55mm
median GA = exp $0.03243 \times [\log_e(250)]2 + 0.001644 \times 55 \times \log_e(250) + 3.813$
= exp (5.300929)
= 200.5 days (equivalent to 28.6 weeks)

Using equations of the median and standard deviation one can easily compute any desired centiles using the relation Pth centile = Median + KSD, where K is the normal equivalent deviate (z score) corresponding to a particular centile, e.g. K = 1.88 for the 97th centile and −1.88 for the 3rd centile. The SDs in this equation are the predicted estimates from the regression analysis.

For example, 3rd centile for GA = exp $(0.05970 * (\log_e HC)^2 + 0.000000006409 * HC^3 + 3.3258) + (-1.88 * (0.6492 * (\text{Median GA} * 0.01)^3 + 2.991))$

Table 3: Selected equations: goodness of fit characteristics of gestational age assessment in days.

Equation ID	1	2
Variables in equation	HC	HC, FL
RMSE (log days)	0.0423	0.0352
R-squared	0.98	0.99
Goodness of fit (%)	5.21	6.20
AIC	16.33	14.69
Variation around the mean of GA estimate (half width of 95% prediction interval (days), at:		
14 weeks	7.1	5.4
16 weeks	7.7	6.5
18 weeks	8.4	7.6
20 weeks	9.4	8.7
22 weeks	10.5	9.8
24 weeks	11.9	10.9
26 weeks	13.5	12.0
28 weeks	15.4	13.2
30 weeks	17.6	14.3
32 weeks	20.1	15.4
34 weeks	23.0	16.5

RMSE = root mean squared error. Goodness of fit = the percentage of predicted estimates of GA outside the 95% prediction interval across all gestational ages. AIC = the Akaike Information Criterion.

Table 4: Prediction of gestational age using the two equations. The median (50th centile) is the most likely gestational age in the absence of pathology. The true gestational age is unlikely to be below the lower limit of the prediction interval (2.5th centile). Data shown in weeks+days

Most likely value (50th centile)	**Lower limit of prediction interval (2.5th centile)**		**Upper limit of prediction interval (97.5th centile)**	
	Equation 1 (HC)	Equation 2 (HC, FL)	Equation 1 (HC)	Equation 2 (HC, FL)
14+0	12+6	13+1	15+0	14+5
14+1	13+0	13+2	15+1	14+6
14+2	13+1	13+3	15+2	15+0
14+3	13+2	13+4	15+3	15+1
14+4	13+3	13+5	15+4	15+2
14+5	13+4	13+6	15+5	15+3
14+6	13+5	14+0	15+6	15+4
15+0	13+6	14+1	16+0	15+5
15+1	14+0	14+1	16+1	16+0
15+2	14+1	14+2	16+2	16+1
15+3	14+2	14+3	16+3	16+2
15+4	14+3	14+4	16+4	16+3
15+5	14+4	14+5	16+5	16+4
15+6	14+5	14+6	16+6	16+5
16+0	14+6	15+0	17+0	16+6
16+1	15+0	15+1	17+1	17+0
16+2	15+1	15+2	17+2	17+1
16+3	15+2	15+3	17+3	17+2
16+4	15+3	15+4	17+4	17+3
16+5	15+4	15+5	17+5	17+4
16+6	15+5	15+5	17+6	17+6

17+0	15+5	15+6	18+1	18+0
17+1	15+6	16+0	18+2	18+1
17+2	16+0	16+1	18+3	18+2
17+3	16+1	16+2	18+4	18+3
17+4	16+2	16+3	18+5	18+4
17+5	16+3	16+4	18+6	18+5
17+6	16+4	16+5	19+0	18+6
18+0	16+5	16+6	19+1	19+0
18+1	16+6	17+0	19+2	19+1
18+2	17+0	17+1	19+3	19+2
18+3	17+1	17+2	19+4	19+3
18+4	17+2	17+3	19+5	19+4
18+5	17+3	17+3	19+6	19+6
18+6	17+4	17+4	20+0	20+0
19+0	17+5	17+5	20+1	20+1
19+1	17+6	17+6	20+2	20+2
19+2	18+0	18+0	20+3	20+3
19+3	18+0	18+1	20+5	20+4
19+4	18+1	18+2	20+6	20+5
19+5	18+2	18+3	21+0	20+6
19+6	18+3	18+4	21+1	21+0
20+0	18+4	18+5	21+2	21+1
20+1	18+5	18+6	21+3	21+2
20+2	18+6	19+0	21+4	21+3
20+3	19+0	19+1	21+5	21+4
20+4	19+1	19+1	21+6	21+6
20+5	19+2	19+2	22+0	22+0
20+6	19+3	19+3	22+1	22+1

21+0	19+4	19+4	22+2	22+2
21+1	19+5	19+5	22+3	22+3
21+2	19+5	19+6	22+5	22+4
21+3	19+6	20+0	22+6	22+5
21+4	20+0	20+1	23+0	22+6
21+5	20+1	20+2	23+1	23+0
21+6	20+2	20+3	23+2	23+1
22+0	20+3	20+4	23+3	23+2
22+1	20+4	20+5	23+4	23+3
22+2	20+5	20+5	23+5	23+5
22+3	20+6	20+6	23+6	23+6
22+4	21+0	21+0	24+0	24+0
22+5	21+1	21+1	24+1	24+1
22+6	21+1	21+2	24+3	24+2
23+0	21+2	21+3	24+4	24+3
23+1	21+3	21+4	24+5	24+4
23+2	21+4	21+5	24+6	24+5
23+3	21+5	21+6	25+0	24+6
23+4	21+6	22+0	25+1	25+0
23+5	22+0	22+1	25+2	25+1
23+6	22+1	22+2	25+3	25+2
24+0	22+2	22+3	25+4	25+3
24+1	22+3	22+3	25+6	25+5
24+2	22+3	22+4	26+0	25+6
24+3	22+4	22+5	26+1	26+0
24+4	22+5	22+6	26+2	26+1
24+5	22+6	23+0	26+3	26+2
24+6	23+0	23+1	26+4	26+3

25+0	23+1	23+2	26+5	26+4
25+1	23+2	23+3	26+6	26+5
25+2	23+3	23+4	27+0	26+6
25+3	23+3	23+5	27+2	27+0
25+4	23+4	23+6	27+3	27+1
25+5	23+5	24+0	27+4	27+2
25+6	23+6	24+1	27+5	27+3
26+0	24+0	24+1	27+6	27+5
26+1	24+1	24+2	28+0	27+6
26+2	24+2	24+3	28+1	28+0
26+3	24+3	24+4	28+2	28+1
26+4	24+3	24+5	28+4	28+2
26+5	24+4	24+6	28+5	28+3
26+6	24+5	25+0	28+6	28+4
27+0	24+6	25+1	29+0	28+5
27+1	25+0	25+2	29+1	28+6
27+2	25+1	25+3	29+2	29+0
27+3	25+2	25+4	29+3	29+1
27+4	25+2	25+5	29+5	29+2
27+5	25+3	25+6	29+6	29+3
27+6	25+4	25+6	30+0	29+5
28+0	25+5	26+0	30+1	29+6
28+1	25+6	26+1	30+2	30+0
28+2	26+0	26+2	30+3	30+1
28+3	26+1	26+3	30+4	30+2
28+4	26+1	26+4	30+6	30+3
28+5	26+2	26+5	31+0	30+4
28+6	26+3	26+6	31+1	30+5

29+0	26+4	27+0	31+2	30+6
29+1	26+5	27+1	31+3	31+0
29+2	26+6	27+2	31+4	31+1
29+3	27+0	27+3	31+5	31+2
29+4	27+0	27+3	32+0	31+4
29+5	27+1	27+4	32+1	31+5
29+6	27+2	27+5	32+2	31+6
30+0	27+3	27+6	32+3	32+0
30+1	27+4	28+0	32+4	32+1
30+2	27+5	28+1	32+5	32+2
30+3	27+5	28+2	33+0	32+3
30+4	27+6	28+3	33+1	32+4
30+5	28+0	28+4	33+2	32+5
30+6	28+1	28+5	33+3	32+6
31+0	28+2	28+6	33+4	33+0
31+1	28+2	29+0	33+6	33+1
31+2	28+3	29+1	34+0	33+2
31+3	28+4	29+1	34+1	33+4
31+4	28+5	29+2	34+2	33+5
31+5	28+6	29+3	34+3	33+6
31+6	29+0	29+4	34+4	34+0
32+0	29+0	29+5	34+6	34+1
32+1	29+1	29+6	35+0	34+2
32+2	29+2	30+0	35+1	34+3
32+3	29+3	30+1	35+2	34+4
32+4	29+4	30+2	35+3	34+5
32+5	29+4	30+3	35+5	34+6
32+6	29+5	30+4	35+6	35+0

33+0	29+6	30+5	36+0	35+1
33+1	30+0	30+6	36+1	35+2
33+2	30+1	30+6	36+2	35+4
33+3	30+1	31+0	36+4	35+5
33+4	30+2	31+1	36+5	35+6
33+5	30+3	31+2	36+6	36+0
33+6	30+4	31+3	37+0	36+1
34+0	30+4	31+4	37+2	36+2
34+1	30+5	31+5	37+3	36+3
34+2	30+6	31+6	37+4	36+4
34+3	31+0	32+0	37+5	36+5
34+4	31+1	32+1	37+6	36+6
34+5	31+1	32+2	38+1	37+0
34+6	31+2	32+3	38+2	37+1
35+0	31+3	32+3	38+3	37+3
35+1	31+4	32+4	38+4	37+4
35+2	31+4	32+5	38+6	37+5
35+3	31+5	32+6	39+0	37+6
35+4	31+6	33+0	39+1	38+0
35+5	32+0	33+1	39+2	38+1
35+6	32+1	33+2	39+3	38+2
36+0	32+1	33+3	39+5	38+3
36+1	32+2	33+4	39+6	38+4
36+2	32+3	33+5	40+0	38+5
36+3	32+4	33+6	40+1	38+6
36+4	32+4	34+0	40+3	39+0
36+5	32+5	34+1	40+4	39+1
36+6	32+6	34+1	40+5	39+3

37+0	33+0	34+2	40+6	39+4

Table 5 Commonly used dating equations and imprecision of gestational age assessment (**half width of** 95% prediction intervals) in days.

Reference	R-squared	12 - 18 weeks	18 - 24 weeks	24 - 30 weeks	30 - 36 weeks
Hadlock (1984)[28]					
BPD	0.967	8.3	12.1	15.3	21.6
HC	0.973	8.3	10.4	14.4	20.9
AC	0.969	11.6	14.4	15.3	20.7
FL	0.971	9.7	12.6	14.6	20.7
HC, BPD	0.974	7.6	10.4	13.9	20.0
HC, FL	0.976	8.4	10.6	13.9	18.8
HC, AC	0.98	7.6	9.4	13.0	17.6
HC, FL, BPD	0.981	7.3	9.5	12.7	17.6
HC, AC, FL	0.981	8.0	10.2	13.2	17.6
HC, BPD, AC, FL	0.981	7. 6	9.8	12.6	17.1
Altman and Chitty (1997)[29]					
HC	NR	8.0	13.0	17.0	22.0

BPD = Biparietal Diameter. HC = Head Circumference. AC = Abdominal Circumference. FL = Femur Length. NR = Not Reported.

Appendix: Details of statistical analytical strategy

Prediction equations were developed using the 'Genetic Algorithm', an iterative, automated machine-learning methodology, whose source code is readily available (Clegg, 2005) http://uk.mathworks.com/matlabcentral/fileexchange/25499-gapolyfitn. Whilst conventional approaches employing fractional polynomial regression are ideally suited for predicting a continuous non-linear relationship using single variables (e.g. HC and GA), these methods do not translate easily to datasets with multiple variables. The purpose of the genetic algorithm was to construct and evaluate a large number of candidate equations, comprising terms based upon subsets of variables from the whole dataset; evaluate their fit of GA_{true} automatically, and to present a single, simplified equation offering the best fit of GA_{true} for a specified number of input variables.

In brief, the process initiated by assembling a generation of 1000 polynomial forms, whose terms comprised random combinations of biometric variables. In accordance with previous work, the powers of terms used to construct equations were limited to 0.5, 1, 2 and 3; with equations subsequently comprising 2, 3 or 4 terms. Individual terms were assembled as random combinations of variables such that a term with power 2 could either comprise the square of a single variable or the product of two separate variables.

The algorithm used a least squares fit to evaluate the coefficients of terms for each equation by minimizing the error between $GA_{predict}$ and GA_{true}. The overall model fit for each equation was assessed using the RMSE, and a generation of equations compared by ranking according to the RMSE. Once ranked, the top 20 percent of equations were selected for modification and iteration within a second round of the algorithm; and the remainder discarded. The genetic algorithm modified the structure of equations using processes designed to mimic natural selection. Whole terms were exchanged between equations (recombination) and the structure of terms was changed at random in 10 percent of equations (point mutation), before the fit of equations was re-assessed, as described elsewhere (Leardi, 1992; Inza 2001; Clegg, 2005; Gutell 2006). When the modification of equations provided a better fit of GA_{true}, equations were conserved and entered a subsequent iteration using the same methods. The search was halted when equation variants failed to improve the RMSE by $>10^{-8}$ (specified arbitrarily). Thus, over many cycles, the genetic algorithm converged upon a single polynomial equation that optimally predicted GA for the specified input variables.

The following data structures were assembled from the INTERGROWTH-21st dataset. Each set of candidate variables (A-J) was processed by the genetic algorithm to determine the multivariable polynomial equation that produced the lowest RMSE for predicting GA for that specific combination of input biometry. It is emphasized that the models increment by the number of terms and the order

with which each variable is entered has no impact upon the selection of terms featuring in the final dataset.

A HC

B HC+logeHCx

C HC+$\log_e$HC+BPD

D HC+$\log_e$HC+BPD+$\log_e$BPD

E HC+$\log_e$HC+BPD+$\log_e$BPD+OFD

F HC+$\log_e$HC+BPD+$\log_e$BPD+OFD+$\log_e$OFD

G HC+$\log_e$HC+BPD+$\log_e$BPD+OFD+$\log_e$OFD+AC

H HC+$\log_e$HC+BPD+$\log_e$BPD+OFD+$\log_e$OFD+AC+$\log_e$AC

I HC+$\log_e$HC+BPD+$\log_e$BPD+OFD+$\log_e$OFD+AC+$\log_e$AC+FL

J HC+$\log_e$HC+BPD+$\log_e$BPD+OFD+$\log_e$OFD+AC+$\log_e$AC+FL+$\log_e$FL

Variables were entered in their natural and natural log state. Generic equations were therefore generated in the form $\log_e GA = aX_b^{\,i} + bX_{b+1}^{\,j} + cX_{b+2}^{\,k}\ldots$, with the limits of X_b restricted to produce equations of 2, 3 or 4 terms. Hence, the structures of the polynomial equations derived were:

$$\log_e GA = aX_1^{\,0-3} + bX_2^{\,0-3} + \text{constant} \quad (1)$$

$$\log_e GA = aX_1^{\,0-3} + bX_2^{\,0-3} + cX_3^{\,0-3} + \text{constant} \quad (2)$$

$$\log_e GA = aX_1^{\,0-3} + bX_2^{\,0-3} + cX_3^{\,0-3} + dX_4^{\,0-3} + \text{constant} \quad (3)$$

where:

a, b, c, d : Coefficients determined by genetic algorithm.

X_1, ... X_4 : Combinations of variables determined by the genetic algorithm. In the case of X^n, X can take the form of a single variable (HC^n) or a combination of variables ($HC^i\,FL^j$) where i+j=n.

Equations of the forms (1) to (3) were constructed for each combination A-J with the terms of each equation restricted to power 2 and subsequently to power 3. Hence, each combination A-J was processed to construct a 2 term, 2 power equation as well as a 2 term, 3 power and a 3 term, 2 power equation, and so on.

The output of the modeling process consisted of a polynomial equation and respective performance indicators including the RMSE, and regression coefficient (r^2). For each equation, absolute residuals between GAtrue – GApredicted were regressed on predicted GA to provide an equation to approximate the SD, and multiplied by the constant $1.96\sqrt{(\pi/2)}$, (Altman 1993) to estimate the 2.5th and 97.5th centiles using the Xrigls function in Stata.

The Genetic Algorithm methodology was tested extensively before running analyses. Specifically, 40 test datasets were assembled within which single and multiple combinations of the constituent variables were replaced by random data generated within the same range as the observed data using the *'runiform'* random number generator in Stata. As an indicator of the robustness of search algorithm, the test datasets were used to determine if any of the random variables would be selected as predictors of GA by the model. This was not the case. Repeated cycles of the whole process were run to assess the reproducibility of results with identical equations resolved each time.

References

R Leardi, R Boggia, and M Terrile. Genetic Algorithms as a Strategy for Feature-Selection. Journal of Chemometrics, 6(5):267–281, 1992.

I Inza, M Merino, P Larranaga, J Quiroga, B Sierra, and M Girala. Feature subset selection by genetic algorithms and estimation of distribution algorithms - A case study in the survival of cirrhotic patients treated with TIPS. Artificial Intelligence in Medicine, 23(2):187–205, 2001.

J Clegg, J F Dawson, S J Porter, and M H Barley. The use of a genetic algorithm to optimize the functional form of a multi-dimensional polynomial fit to experimental data. Evolutionary Computation, The 2005 IEEE Congress on,(1):928 – 934, 2005

R R Gutell and R K Jansen. Thesis: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion, 2006.