Accepted: 13 July 2024

DOI: 10.1111/1471-0528.17922

BJOG An International Journal of Obstetrics and Gynaecology

RESEARCH ARTICLE

External validation of models to estimate gestational age in the second and third trimester using ultrasound: A prospective multicentre observational study

Alice Self[1] | **Michael Schlussel**[2] | **Gary S. Collins**[2] | **Ferdinand Dhombres**[3,4] | **Nicolas Fries**[4] | **Georges Haddad**[4,5] | **Laurent J. Salomon**[4,6] | **Mona Massoud**[4,7] | **Aris T. Papageorghiou**[1,8]

[1]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK

[2]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

[3]Armand Trousseau University Hospital, Sorbonne University, Paris, France

[4]Collège Francais d'Échographie Foetale, Paris, France

[5]Simone Veil Hospital, Blois, France

[6]Maternité, Hopital Necker Enfants Malades, Université Paris Descartes, Paris, France

[7]Obstetrics and Fetal Medicine Unit, Hôpital Lyon Sud, Hospices Civils de Lyon and FLUID Team, Lyon Neurosciences Research Center, INSERM U1028, CNRS UMR5292, Lyon-1 University, Bron, France

[8]Oxford Maternal and Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK

Correspondence
Aris T. Papageorghiou, Nuffield Department of Women's and Reproductive Health, University of Oxford, The Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU, UK.
Email: aris.papageorghiou@wrh.ox.ac.uk

Abstract

Objectives: Accurate assessment of gestational age (GA) is important at both individual and population levels. The most accurate way to estimate GA in women who book late in pregnancy is unknown. The aim of this study was to externally validate the accuracy of equations for GA estimation in late pregnancy and to identify the best equation for estimating GA in women who do not receive an ultrasound scan until the second or third trimester.

Design: This was a prospective, observational cross-sectional study.

Setting: 57 prenatal care centres, France.

Participants: Women with a singleton pregnancy and a previous 11–14-week dating scan that gave the observed GA were recruited over an 8-week period. They underwent a standardised ultrasound examination at one time point during the pregnancy (15–43 weeks), measuring 12 foetal biometric parameters that have previously been identified as useful for GA estimation.

Main Outcome Measures: A total of 189 equations that estimate GA based on foetal biometry were examined and compared with GA estimation based on foetal CRL. Comparisons between the observed GA and the estimated GA were made using R^2, calibration slope and intercept. RMSE, mean difference and 95% range of error were also calculated.

Results: A total of 2741 pregnant women were examined. After exclusions, 2339 participants were included. In the 20 best performing equations, the intercept ranged from −0.22 to 0.30, the calibration slope from 0.96 to 1.03 and the RSME from 0.67 to 0.87. Overall, multiparameter models outperformed single-parameter models. Both

Mona Massoud and Aris T. Papageorghiou are joint last authors.

All authors have completed the Unified Competing Interest form (available upon request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

the 95% range of error and mean difference increased with gestation. Commonly used models based on measurement of the head circumference alone were not amongst the best performing models and were associated with higher 95% error and mean difference.

Conclusions: We provide strong evidence that GA-specific equations based on multiparameter models should be used to estimate GA in late pregnancy. However, as all methods of GA assessment in late pregnancy are associated with large prediction intervals, efforts to improve access to early antenatal ultrasound must remain a priority.

Trial Registration: The proposal for this study and the corresponding methodological review was registered on PROSPERO international register of systematic reviews (registration number: CRD4201913776).

KEYWORDS

biometry, due date, gestational age, growth, post-term, pregnancy, pregnancy dating, preterm, screening, ultrasound dating

1 | INTRODUCTION

Accurate gestational age (GA) assessment is a key component of obstetric care for timing interventions, such as the administration of steroids to the mother with preterm labour to enhance neonatal lung maturity; management of prematurity and foetal growth restriction; and induction of labour for obstetric complications and postmaturity. It is also important at a population level to estimate rates of preterm birth and of small and large-for-gestational age (SGA and LGA) foetuses, which are important phenotypic distinctions. Ultrasound measurement of foetal crown–rump length (CRL) between 45 and 84 mm is recognised as the most accurate way to estimate GA, but due to foetal curling cannot be applied beyond the first trimester.[1]

Nevertheless, early ultrasound remains limited in many regions of the world,[2] and a high proportion of women in low- and middle-income countries (LMICs) book after 20 weeks.[3–5] Even in well-resourced settings, over 5% of women receive no antenatal care until the late second or third trimester.[6,7] These women are amongst the most vulnerable,[8] added to which, adverse outcomes have been found to be higher amongst those with uncertain last menstruation.[9]

The WHO recommends that all women should have an ultrasound scan before 24 weeks, in part to accurately date the pregnancy.[10] However, many women are still scanned in late pregnancy, especially those uncertain of their last menstruation. Here, there is no agreed way on how to best estimate GA: Many pregnancies will be assigned a GA estimated from head circumference (HC) alone, but imprecision of ±14.4 to 17.0 days at 24–30 weeks and ±20.9 to 22.0 days at 30–36 weeks has been reported for the most commonly used HC equations.[11] A number of articles and guidelines have suggested that composite equations using multiple parameters might be preferable as they are more accurate than a single parameter alone in the second and third trimesters,[12–14] but independent comparison is lacking.

We have previously published a systematic review of the methodological quality of studies reporting equations for estimating GA in the second and third trimesters.[1] We showed that estimation by maternal fundal height measurement is inferior to ultrasound. We also identified the 10 most methodologically robust studies, based on study quality. However, a head-to-head comparison to ascertain which equations best estimated GA was not possible without an independent data set that measured all the different variables included in all these equations. We required a novel data set and so we undertook a large prospective study and performed comparison of accuracy of candidate equations estimating GA that we identified previously,[1] in order to make recommendations for clinical practice.

2 | METHODS

2.1 | Sonographer recruitment

This was a prospective, observational, cross-sectional study assessing the precision of ultrasound biometry in predicting GA. We used existing 'flash' methodology,[15–17] a pragmatic, short and focussed type of study that is embedded in routine clinical practice. 'Flash' studies are undertaken using the countrywide network of sonographers who are members of the French College of Foetal Ultrasound (Collège Français d'Échographie Fœtale [CFEF]) and require only minor additional steps to routine practice.

We invited sonographers to download the study protocol and to take an online training course (www.cfef.org) that reviewed the aims of the study, inclusion criteria, methodology for taking the measurements and quality control criteria of each view (File S1). Only sonographers who completed the course and sent five complete and high-quality image-sets (of the requisite eight views) were eligible to participate in the study.

2.2 | Recruitment of women

Pregnant women with a singleton pregnancy without an obvious congenital abnormality were prospectively included over a fixed study period of 8 weeks. Scans were undertaken routinely at 11–14, 21–24 or 30–34 weeks' gestation according to standard antenatal care in France, or in between based on medical history. Maternal age, parity, body mass index (BMI), date of the first-trimester ultrasound and the recorded CRL measurement were collected. For all included women, pregnancy dating was based on CRL measurement in the first trimester, as recommended by the French College of Obstetrics and Gynecology (CNGOF)[18] using the Robinson and Flemming equation.[19] This was the observed GA that was subsequently used as the 'ground truth' for testing later pregnancy dating equations. Although women were subsequently scanned more than once, each woman was only scanned once during the study period.

In order to gather data to test the different GA models, a total of 12 ultrasound measurements, yielding 13 different biometric parameters to be tested, were obtained:

1. Foetal head circumference (HC);
2. Biparietal diameter measured by placing callipers in an outer to outer (BPDoo) position;
3. Biparietal diameter measured by placing callipers in an outer to inner (BPDoi) position;
4. Occipital frontal diameter (OFD);
5. Abdominal circumference (AC);
6. External interorbital distance (EIOD);
7. Transverse cerebellar diameter (TCD);
8. Femur length (FL);
9. Tibia length (TL);
10. Foot length (FtL);
11. Right kidney length (RKL);
12. Left kidney length (LKL); and
13. Mean kidney length (MKL) was calculated from the mean of the RKL and LKL when both measurements were available.

These measurements were the most frequently measured foetal parameters identified in our previous systematic review.[1] Manual data entry was undertaken and data cleaning and identification of outliers performed. Individual sonographers were contacted to verify locally stored images and correct any errors. We excluded women and their pregnancies if there was no record of CRL measurement during the first trimester and all foetuses with congenital malformations, or intrauterine growth restriction. Additional information about exclusions and a complete list of values deemed as outliers is presented in File S2.

2.3 | Identification of equations

In a previous systematic review,[1] we assessed the methodology and risk of bias of 97 publications describing studies proposing models for GA estimation using ultrasound biometric parameters. These publications described 284 equations, of which 237 were composed of biometric parameters considered feasible to be measured reproducibly in this study in a routine setting. Of the 237 models, 48 were excluded[20–31] as they did not produce biologically plausible GA estimates; this left 189 unique models from 73 publications[11,23,24,26–28,31–97] for testing (File S3).

2.4 | Statistical analysis

When considering the accuracy of GA estimation, we need to assess the mean difference between observed and predicted GA (representing bias); as well as the range of error between observed and predicted GA (representing in some sense the random error). The former is more population-specific, while the latter a function of variation, such as biological and measurement variability and variation due to the underlying equation linking the biometric variable(s) to GA. Therefore, our overall statistical strategy was to establish a number of metrics related to these key concepts.

We applied each GA estimation model to each participant in the final dataset. Thus, for each participant, an estimated GA (predicted GA) was calculated for each included model based on their respective biometric variables. The observed GA (reference standard) was based on foetal CRL previously measured at 11–14 weeks.[19]

The performance of each model for predicting GA was assessed based on the coefficient of determination (R^2) and represents the proportion of the variance in GA explained by the included predictors in any given model (a perfect model would explain 100% of the GA variance, so the R^2 ranges from 0 to 1). The models' calibration slope and calibration-in-the-large (intercept) are estimated by regressing the observed on predicted values of GA. The calibration slope evaluates the spread of the estimated values and has a target value of 1. A slope <1 suggests that predicted values are too high for individuals with low observed values and too low for those with high observed values. A slope >1 suggests the opposite. The calibration-in-the-large has a target value of 0; negative values suggest overestimation, whereas positive values suggest underestimation. The root mean square error (RMSE) was calculated to reflect the average difference between the predicted and observed values of GA, with a value of zero indicating no difference between the predicted and the true GA.

Models that consistently presented better performance across all four performance parameters were selected. To do this, we ranked models according to the best estimate obtained for the performance parameters described above (R^2 calibration slope, intercept, and RMSE). We then simply summed each model's rank for these four parameters and selected the 20 with lowest scores for detailed reporting. We assessed the models' overall calibration graphically in terms of GA by plotting agreement between predicted and observed values.

It is clinically beneficial that GA estimation models work throughout gestation. However, some of the validated

models were developed for specific ranges of GA. We assessed all models on both the full data set and limited to the GA range they were developed for.

To aid clinical interpretation, we also present a 95% range of error, estimated as the interval defined by the 2.5th and 97.5th percentiles of the difference between observed and predicted GAs for all models at 18–23^{+6}; 24–29^{+6}; 30–35^{+6}; and 36–40^{+0} weeks of gestation. The mean difference between observed and predicted GA for each interval was estimated to provide an indication of over or underestimation. The narrower the 95% range of error is, and the closer to zero the mean difference between observed and predicted GAs is, the more accurate is the predicted GA provided by any given model.

Women with incomplete or outlier values (File S2) for any of the biometric parameters of interest were excluded, and a complete case analysis carried out.

2.5 | Sample size

The smallest R^2 value reported in our previous systematic review was 0.7.[1] Assuming the models would perform worse in the validation data set (so, anticipating an R^2 upon the external validation data of 0.5), and targeting a 95% CI with a narrow width of 0.1, the minimum sample size required for external validation would be approximately 770 individuals.[98] In order to achieve this in a fixed period of 8 weeks, we anticipated requiring 40–80 centres, depending on individual throughput. As we anticipated that not all measurements would be feasible in all women at all gestational ages, and due to uncertainties related to the COVID-19 pandemic, we took a decision to oversample, and the analysis of 2339 women from the data set clearly exceeds the minimum sample size requirement to evaluate model performance.

TABLE 1 Summary characteristics of complete cases with no outliers, complete cases and entire study population.

	Complete cases with no outliers					Complete cases including outliers					Entire population				
Variable	***n***	**Mean**	**SD**	**Min**	**Max**	***n***	**Mean**	**SD**	**Min**	**Max**	***n***	**Mean**	**SD**	**Min**	**Max**
Age (years)	2295	32.0	5.0	16.0	50.9	2406	32.0	5.0	16.0	50.9	2692	32.0	5.0	16.0	50.9
Parity	2339	1[a]	–	0	10	2451	1[a]	–	0	10	2741	1[a]	–	0	10
0	1018	43.52[b]	–	–	–	1055	43.04[b]	–	–	–	1192	43.49[b]	–	–	–
1	837	35.78[b]	–	–	–	883	36.03[b]	–	–	–	985	35.94[b]	–	–	–
2	328	14.02[b]	–	–	–	344	14.04[b]	–	–	–	379	13.83[b]	–	–	–
3 or more	156	6.67[b]	–	–	–	169	6.9[b]	–	–	–	185	6.75[b]	–	–	–
Weight (kg)	2331	70.2	13.9	40.0	150.0	2443	70.2	13.9	40.0	150.0	2729	69.9	13.7	40.0	150.0
Height (cm)	2324	164.5	6.3	123.0	185.0	2435	164.5	6.3	123.0	185.0	2724	164.5	6.2	123.0	185.0
BMI (kg/m^2)	2318	25.9	4.9	14.7	67.6	2429	25.9	4.9	14.7	67.6	2715	25.8	4.8	14.7	67.6
IGGA (weeks)	2339	27.1	5.1	15.9	38.9	2451	27.1	5.2	15.9	43.1	2741	27.2	5.2	15.9	43.1
RobGA2 (weeks)	2339	26.9	5.1	15.7	38.7	2451	26.9	5.2	15.7	42.9	2741	27.0	5.2	15.7	42.9
HC (mm)	2339	244.4	50.2	119.9	361.0	2451	244.1	50.4	81.0	361.0	2735	245.4	51.0	47.0	361.0
AC (mm)	2339	228.4	55.3	105.1	367.3	2451	228.1	55.4	58.0	367.3	2735	229.3	55.7	58.0	367.3
FL (mm)	2339	49.3	12.0	18.6	73.0	2451	49.2	12.0	18.6	73.0	2738	49.5	12.1	18.6	77.7
BPDoo (mm)	2339	68.1	14.7	32.9	98.6	2451	68.1	14.7	32.9	98.6	2739	68.5	14.8	32.9	98.6
BPDio (mm)	2339	66.2	14.6	32.3	96.2	2451	66.2	15.3	21.0	294.5	2661	66.6	15.3	21.0	294.5
OFD (mm)	2339	86.0	17.2	42.6	122.3	2451	86.3	19.6	33.0	405.8	2624	86.6	19.5	33.0	405.8
TCD (mm)	2339	31.7	9.1	14.7	54.9	2451	31.7	9.1	14.7	59.5	2709	32.0	9.2	14.7	59.5
BND (mm)	2339	41.6	7.5	22.7	67.4	2451	41.5	8.2	6.1	161.6	2677	41.6	8.2	6.1	161.6
RKL (mm)	2339	31.2	7.9	12.6	53.4	2451	31.2	7.9	11.4	53.4	2699	31.4	8.0	2.7	55.3
LKL (mm)	2339	31.3	7.9	13.5	53.0	2451	31.3	7.9	6.6	53.0	2688	31.4	8.0	4.1	53.0
MKL (mm)[c]	2339	31.3	7.8	13.2	51.5	2451	31.2	7.8	11.3	51.5	2686	31.4	7.9	11.3	53.6
FtL (mm)	2339	51.8	13.6	19.1	83.6	2451	51.8	13.6	19.1	83.6	2706	52.0	13.7	19.1	83.6
TL (mm)	2339	43.5	10.7	16.2	66.1	2451	43.4	10.7	16.2	81.9	2679	43.6	10.7	16.2	81.9

Abbreviations: AC, abdominal circumference; BPDoi, biparietal diameter outer to inner; BPDoo, biparietal diameter outer to outer; EIOD, external interorbital distance; FL, femur length; FtL, foot length; HC, head circumference; LKL, left kidney length; MKL, mean kidney length; OFD, occipital frontal diameter; RKL, right kidney length; TCD, transverse cerebellar diameter; TL, tibia length.

[a]Median.

[b]Percentage.

[c]MKL as estimated based on RKL and LKL for women with both measures only.

BJOG An International Journal of Obstetrics and Gynaecology

2.6 | Ethics and patient and public involvement

There was no patient or public involvement in the development of the study design or the way it was conducted. Patients were first involved in the research when provided with information about the data collection at their routine scans.

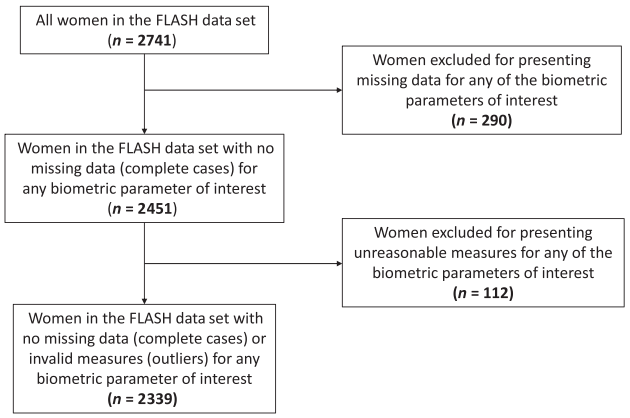

FIGURE 1 Flow diagram of study participants.

3 | RESULTS

Between 6 September 2021 and 29 October 2021, 124 sonographers agreed to participate of whom 57, from 57 centres, fulfilled the inclusion criteria of the study. During the study period, the sonographers performed scans on a total of 2741 women. Over 97% of individual biometry measurements were successfully attained at recruitment (data available upon request). As we only included data sets in the analysis where all 12 measurements of all structures were obtained, we excluded 290 (10%) women with incomplete data of one or more measurement needed to compute the GA equations. After excluding 112 (4%) outliers, this left a final sample of 2339 women. The same number of scans, with complete sets of data across all parameters between 15^{+5} and 38^{+6} weeks of gestation, were included for analysis (Table 1). Figure 1 depicts the flow of participants through the study and numbers excluded. File S4 shows the number of scans included per week of gestation.

The performance of all 189 GA equations was then estimated (Files S5 and S6). We examined the 20 equations ranked as best performing (Table 2, and their calibration plots in File S7). The intercept ranged from −0.219 to 0.297 amongst the top 20. Values of R^2 demonstrated that a high proportion of the variation in the CRL-calculated GA that was explained by the models estimating GA with

TABLE 2 Performance of the 20 best performing models assessed on all complete sets of data between 18 and 40 weeks' gestation.

Model number	Authors	Year	Parameters used	GA range developed for (weeks)	RoB score[a]	Intercept	R^2	Slope	RMSE
37	Dare et al.[40]	2004	FL	Unknown	24%	0.04	0.97	1.02	0.87
82	Hadlock et al.[53]	1984	FL	14–42	28%	0.08	0.97	1.02	0.86
85	Hadlock et al.[53]	1984	BPDoi and FL	14–42	28%	0.14	0.98	0.98	0.76
88	Hadlock et al.[53]	1984	AC and FL	14–42	28%	−0.02	0.98	1.01	0.71
90	Hadlock et al.[53]	1984	BPDoi, AC and FL	14–42	28%	0.02	0.98	1.00	0.68
92	Hadlock et al.[53]	1984	HC, AC and FL	14–42	28%	0.30	0.98	1.00	0.75
93	Hadlock et al.[53]	1984	BPDoi, HC, AC and FL	14–42	28%	0.21	0.98	0.99	0.71
99	Hill et al.[24]	1992	FL	13–43	10%	−0.04	0.97	1.03	0.86
100	Hill et al.[24]	1992	BPDoi and AC	13–43	10%	−0.02	0.97	0.97	0.86
102	Hill et al.[24]	1992	BPDoi and FL	13–43	10%	0.11	0.98	1.01	0.74
105	Hill et al.[24]	1992	AC and FL	13–43	10%	0.09	0.97	1.00	0.83
106	Hill et al.[24]	1992	BPDoi, HC and AC	13–43	10%	−0.03	0.98	0.96	0.83
107	Hill et al.[24]	1992	BPDoi, AC and FL	13–43	10%	0.01	0.98	1.01	0.69
109	Hill et al.[24]	1992	HC, AC and FL	13–43	10%	0.24	0.98	0.99	0.73
155	AMANHI Group[66]	2020	TCD and FL	24–36 + 6	66%	−0.13	0.98	0.99	0.68
157	AMANHI Group[66]	2020	TCD and AC	24–36 + 6	66%	−0.20	0.98	1.02	0.76
158	AMANHI Group[66]	2020	BPDoi, TCD, AC and FL	24–36 + 6	66%	−0.22	0.99	0.97	0.67
199	Rodriguez-Sibaja et al.[79]	2020	TCD	14–42	79%	0.12	0.98	1.00	0.82
203	Satish Prasad, Likhitha[81]	2014	FL	15–40	21%	0.04	0.97	0.98	0.86
229	Varol et al.[31]	2001	FL	13–40	21%	0.14	0.97	1.00	0.85

[a]Risk of Bias score from "Second and third trimester estimation of gestational age using ultrasound or maternal symphysis-fundal height measurements: A systematic review" [13].

all exceeding 0.972. Values for the top 20 models ranged from 0.961 to 1.002 for calibration slope and RSME values ranged between 0.67 and 0.87. In general, multiparameter models outperformed single parameter models. There were no HC-only models amongst the best performing 20; rather combinations of HC, BPD, TCD, AC and FL performed best. Thus, models based on BPDoi, AC and FL models[24,53] consistently ranked most highly across all four measures of performance followed by an AC and FL model[53] and a TCD and FL model.[66] However, all the 20 top performing models had good performance overall.

Clinically, the 95% range of prediction (from the 2.5th to the 97.5th centile of actual gestation) is a useful indicator of model accuracy. In Table 3, we summarise the 95% range of error for the 20 models. Results for all 189 models can be found in File S6. Amongst the top 20 models, 95% of all estimates fell within ±7 days of the observed GA at 18^{+0}–23^{+6} weeks. This was very similar to commonly used HC models.[33,50,51,53] As gestation increased, the 95% range of error also increased; thus, between 24^{+0} and 29^{+6} weeks, the more accurate models had 95% of estimates within 9.5 days of the observed GA, which increased to between ±10.9 and 15.2 days at 30^{+0}–35^{+6} weeks. In comparison, the lowest 95% range of error for an HC-only model was ±18.6 days.[33]

Beyond 36 weeks, GA estimation was generally very poor, but only 20 women were eligible for this analysis. In 155 (79%) models, the GA was underestimated in the 36–40^{+0}-week category. The 95% range of error was 3 weeks or more for 132 (67%) of models. Even amongst the 20 best-performing models, the 95% range of error varied between ±13–28 days and was ±27–31 days for HC-only models.

In subgroup analysis, limiting to GA ranges that the equations were designed for, three models developed by the AMANHI group no longer featured in the top 20[66](see File S8). Instead, two additional multi-parameter models[53] and a TCD only model[39] were included.

The mean difference was consistently low until 36–40^{+0} weeks. At 18–23^{+6} and 24 and 27^{+6} weeks, the mean difference was predominantly within 2 days of the observed GA for many equations. The mean difference of models ranged from 2.6 to −1.4 days at 30^{+0}–35^{+6} weeks, and this compared favourably to commonly used HC models (1.1 to −5.5 days). Between 36 and 40^{+0} weeks, the best models tended to underestimate GA by 7–9 days; this underestimation was even more notable in HC-only models (11–16 days).[33,50,51]

TABLE 3 95% range of error and mean difference (days) for the twenty best performing models assessed on all complete sets of data between 18 and 40 weeks' gestation.

Model number	Authors	Year	Parameters used	GA range developed for (weeks)	18–23 + 6 weeks (n = 1141)			
					p2.5	p97.5	95% range of error	Mean difference
37	Dare et al.[40]	2004	FL	Unknown	−7.1	11.6	9.3	1.5
82	Hadlock et al.[53]	1984	FL	14–42	−6.5	10.1	8.3	0.9
85	Hadlock et al.[53]	1984	BPDoi and FL	14–42	−7.8	5.7	6.7	−1.1
88	Hadlock et al.[53]	1984	AC and FL	14–42	−5.2	7.8	6.4	0.9
90	Hadlock et al.[53]	1984	BPDoi, AC and FL	14–42	−5.8	6.4	6.5	0.1
92	Hadlock et al.[53]	1984	HC, AC and FL	14–42	−7.5	4.2	6.1	−1.9
93	Hadlock et al.[53]	1984	BPDoi, HC, AC and FL	14–42	−7.3	4.4	5.8	−1.4
99	Hill et al.[24]	1992	FL	13–43	−5.4	11.5	5.9	2.0
100	Hill et al.[24]	1992	BPDoi and AC	13–43	−7.5	7.4	8.5	−0.2
102	Hill et al.[24]	1992	BPDoi and FL	13–43	−6.7	7.8	7.4	0.0
105	Hill et al[24]	1992	AC and FL	13–43	−6.8	7.7	7.3	0.1
106	Hill et al.[24]	1992	BPDoi, HC and AC	13–43	−7.0	6.3	7.3	−0.6
107	Hill et al.[24]	1992	BPDoi, AC and FL	13–43	−5.6	8.1	6.6	0.8
109	Hill et al.[24]	1992	HC, AC and FL	13–43	−7.3	4.4	6.9	−1.6
155	AMANHI Group[66]	2020	TCD and FL	24–36 + 6	−5.7	8.3	5.9	1.2
157	AMANHI Group[66]	2020	TCD and AC	24–36 + 6	−5.2	9.4	7.0	2.4
158	AMANHI Group[66]	2020	BPDoi, TCD, AC and FL	24–36 + 6	−5.4	7.5	7.3	1.0
199	Rodriguez-Sibaja et al.[79]	2020	TCD	14–42	−10.1	8.4	6.5	−0.4
203	Satish Prasad, Likhitha[81]	2014	FL	15–40	−7.6	9.3	9.3	−0.2
229	Varol et al.[31]	2001	FL	13–40	−7.6	9.3	8.4	−0.1

Note: p2.5 and p97.5 are the 2.5th and 97.5th centiles of the difference distribution; 95% range of error is the half width of the range of error of 95% of all GA estimates. All units in days.

4 | DISCUSSION

4.1 | Principal findings

Using a prospective, multi-centre, national observational study we developed data set of 2339 foetuses with a complete set of 12 different ultrasound measurements used in models for GA estimation. This has enabled the identification of the 20 best performing models (from 8 studies) for estimation of GA from 189 equations using one or more of 13 biometric parameters.

Our analysis demonstrates that, although there is no single equation that stands out, multiparameter models provided better estimates of GA. We believe this constitutes strong evidence against the use of single-parameter models based on HC for GA estimation in women who book late in pregnancy. The best performing models consistently contained varying combinations of HC, BPD, TCD, AC and FL (Table 2).

We describe the clinical relevance of these findings by showing the range in which 95% of the GA estimates for each model fall. Until 29^{+6} weeks' gestation, commonly used HC-only equations[33,50,51,53] performed similarly, although less well than the best performing models. However, between 30 and 35^{+6} weeks, the half-width 95% range of error was over a week more than for the best performing models.

4.2 | Limitations and strengths and comparison with other studies

A major strength of this FLASH study is that real-world data were obtained spanning the second and third trimesters and include complete data on the 13 biometric parameters to be tested. Data were obtained by experienced sonographers that undertake daily practice in France, and quality control by prestudy submission of images was undertaken. There are some disadvantages to this approach: First, the distribution of data points across gestation were influenced by routine practice and so were not evenly spread. We do not believe this impacts the credibility of our findings due to the overall size of the study. Second, it was implausible to reproducibly measure all 25 parameters that have been used for GA estimation,[1] and models using placental thickness and sacral length were not tested. However, over 83% equations identified in previous work[1] were tested, including all models from ultrasound studies reporting sound methodology. Although

24–29 + 6 weeks (*n* = 165)				30–35 + 6 weeks (*n* = 974)				36–40 weeks (*n* = 20)			
p2.5	p97.5	95% range of error	Mean difference	p2.5	p97.5	95% range of error	Mean difference	p2.5	p97.5	95% range of error	Mean difference
−7.6	15.9	11.7	2.3	−13.9	11.3	12.6	−2.2	−38.9	−3.3	17.8	−15.1
−9.5	14.0	11.8	0.6	−14.7	12.6	13.7	−2.1	−39.4	0.8	20.1	−12.5
−9.3	9.8	9.5	−0.3	−13.4	12.2	12.8	−0.8	−37.7	5.1	21.4	−10.9
−7.8	13.2	9.6	0.6	−12.3	12.5	13.9	−0.7	−27.8	6.0	22.0	−9.4
−8.0	11.1	10.5	0.1	−12.0	12.0	12.4	−0.3	−30.9	5.9	16.9	−8.9
−9.1	8.6	9.6	−1.3	−12.5	10.0	12.0	−2.3	−36.1	7.3	18.4	−12.2
−8.4	8.0	8.9	−0.9	−12.2	10.1	11.3	−1.5	−34.9	6.1	21.7	−11.5
−8.2	15.3	8.2	1.9	−13.8	12.6	11.2	−1.7	−38.8	−0.3	20.5	−13.0
−9.2	12.1	11.8	0.2	−13.6	16.7	13.2	0.6	−34.7	10.4	19.2	−7.6
−7.8	10.9	10.6	0.6	−13.4	10.6	15.2	−1.6	−38.5	0.7	22.5	−12.9
−10.4	12.6	9.4	−1.4	−14.5	14.5	12.0	−1.5	−29.6	10.9	19.6	−7.1
−8.0	11.7	11.5	1.1	−11.9	17.0	14.5	1.1	−32.1	13.6	20.3	−9.1
−7.8	12.7	9.8	1.0	−12.5	11.1	14.4	−1.0	−32.7	2.2	22.9	−11.3
−8.3	8.6	10.2	−0.6	−11.9	10.5	11.8	−1.8	−32.3	8.2	17.4	−11.5
−7.5	14.7	8.5	1.9	−10.6	11.2	11.2	0.8	−26.0	5.3	20.2	−9.1
−8.1	10.2	11.1	1.3	−13.2	12.2	10.9	0.3	−28.4	11.5	15.6	−8.3
−6.1	13.6	9.2	2.7	−8.3	12.0	12.7	2.3	−26.8	5.2	19.9	−7.7
−12.5	12.9	9.9	0.4	−15.9	10.0	10.1	−1.3	−29.5	−3.6	16.0	−16.4
−9.5	14.1	12.7	0.4	−13.6	15.2	13.0	−0.3	−38.0	5.0	12.9	−9.2
−9.6	14.0	11.8	0.3	−14.5	12.7	14.4	−2.0	−39.2	0.0	21.5	−12.8

few data points were available from 36 weeks, many studies show poor accuracy of GA estimation in very late pregnancy, meaning caution is recommended when estimating GA.[1,11] It is clearly the case that GA estimation based on biometry will underestimate GA when a foetus is small and overestimate GA when the foetus is large, and as effects of growth aberrations become more common and pronounced, the error of GA estimation will be significant in late pregnancy. Third, the design of a FLASH study meant we were unable to obtain outcome data and ascertain the prevalence of SGA within our cohort. However, the comparator, or 'gold standard' of gestational age was based on previous CRL, and we do not believe that a lack of outcome data should bias these results.

Despite these limitations, we have demonstrated the feasibility of FLASH methodology by successfully recording measurements from 2741 women over an eight-week time period, which would usually have taken many months. This enabled us to compare over 200 different equations head-to-head. For the first time, it has been possible to externally validate equations containing kidney length, external interorbital diameter, tibia length and foot length against more routinely measured parameters.

To our knowledge, this work represents the first external validation of such a comprehensive number GA estimation models. It has generated a data set upon which future models can be validated and compared (data available on reasonable request).

Many of the studies identified in our systematic review reported R^2 as the only marker of good fit. The R^2 was excellent for many of the models we tested, demonstrating the correlation of foetal size and GA. It is important to highlight that no single metric can identify the most accurate model for GA estimation; thus, detailed consideration by our team resulted in the four statistical measures used to describe the fit of a models. While the 95% prediction interval is a clinically useful descriptor to understand the expected range of error, it is not a measure of model performance. In this paper, the 95% range of error conveys the observed error on our data set as opposed to giving a prediction of what error is to be expected. A true 95% prediction interval can only be calculated at the time of model development on the original data from which it was developed.

All models that estimate GA based on biometry assume that foetal size equates to foetal age. This is problematic for four reasons: It requires accurate measurement; ignores biological variation; results in underestimation of GA in small babies (and overestimation in large babies); and is associated with wide range of error as growth rates of many structures flatten out in late gestation. Using single parameters to form an estimate may exacerbate all of these. Thus, the optimal method of estimation would be to identify a parameter that is unaffected by foetal size. One potential measure is TCD, which is measurable even in late gestation[66] and appears to be less impacted by extremes of foetal growth than other parameters,[99] likely due to brain-sparing noted with growth restriction.[100] In our data set, TCD-only models did not consistently produce more accurate estimates of GA than other multiparameter models; however, they have the advantage of needing a single, simple parameter, and this may be useful in lower resource settings when training health-workers who would not normally undertake ultrasound.[101] Interestingly, between 36 and 40 weeks, TCD-only models[34,81] produced some of the lowest 95% ranges of error. However, these data need to be interpreted with caution owing to the low number of babies included at this gestation. Future work should test the performance of the top performing models, including those of TCD (+/– other parameters) for GA estimation in a cohort including SGA babies.

The exhaustive analysis presented suggests that it is unlikely that any new 2D ultrasound measurements will further improve estimates of GA, as variation in foetal size will remain a limiting factor.[13] Thus, efforts to develop new equations may be better spent developing alternative methods of GA estimation. Amniotic fluid volume,[102] bowel maturation patterns,[103] changes in epiphyses and ossification centres[104] placental maturation[105] and cerebellar appearance[106,107] have all been postulated as ways to refine error in estimating GA, but may not be exempt from changes in SGA foetuses.[104] Promising recent work has shown that there may be information relating to GA in 2D ultrasound images not easily recognised by the human eye. Image-based machine-leaning can use this information to estimate GA to within a few days of the true GA, even in the presence of growth abnormalities.[108,109] Machine-learning models of data can also estimate GA with accuracy 3–5 times greater than previous biometry models by analysing trajectories over time.[110]

4.3 | Interpretation and clinical implications

Many women in LMIC settings still present late in pregnancy due to limitations in antenatal care and ultrasound. The proportion of babies that are SGA is particularly high in underserved regions, and the accuracy of models to estimate GA in late pregnancy will be affected by this. In this study, we excluded growth restricted foetuses and cannot comment on model performance within SGA subpopulations. However, even if there are differences in the mean GA estimation in different populations, our findings are very valuable to estimate the range of error associated with different models. All this suggests that, while identifying and using the most accurate biometry-based models is important, our efforts must be focussed on improving access to early antenatal care.

5 | CONCLUSION

Using a unique data set from a prospective multicentre study, we were able to undertake assessment of the accuracy of most commonly available GA estimation models. Late GA estimation using foetal biometry remains less

accurate than first-trimester measurement of CRL, and the later the assessment is made, the greater the error. Thus, efforts to engage pregnant women in antenatal care early must continue. We present strong evidence that multiparameter models should replace commonly used single parameter HC-based models for the estimation of GA. Models including TCD perform well, are better in very late pregnancy and may be more appropriate for use in settings of high SGA burden.

AUTHOR CONTRIBUTIONS

AS, MS, GSC, FD, NF, GH, LS, MM and ATP were all contributed to the conceptualisation and design of the study. AS, MS, GSC and ATP analysed and interpreted data. Original drafting was done by AS, MS, MM and ATP with revision contributions and approval of the final version from GSC, FD, NF, GH and LS. Supervision came from GSC, MM and ATP. The lead author (the manuscript's guarantor—Alice Self) affirms that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ACKNOWLEDGEMENTS

We are indebted to Philippe Billen for his work designing and managing the website for data collection. We would also like to thank Dr Lama Daher and Dr Christos Ioannou both from the Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK, at the time of the articles conception, and Nia Roberts from Bodleian Health Care Libraries, University of Oxford, Oxford, UK who all made an invaluable contribution to identifying and reviewing the studies included in our literature review. The equations validated in this study all came from that review.[1]

FUNDING INFORMATION

Aris T. Papageorghiou is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC).

CONFLICT OF INTEREST STATEMENT

Aris T. Papageorghiou is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the NIHR Biomedical Research Centre funding scheme. He is co-founder and consults for the spin out company Intelligent Ultrasound via Oxford University Innovations. He is the Editor in Chief of BJOG but had no role in handling of this manuscript. There are no other conflicts of interest.

DATA AVAILABILITY STATEMENT

Our data can be made available, upon request to the corresponding author, for validation of other models or for other analyses after receipt of a detailed protocol for the study proposed.

ETHICS APPROVAL

The study was approved by the Ethical Committee of the 'Hospices Civils de Lyon' with the number 417/69HCL21_0565.

ORCID

Alice Self https://orcid.org/0000-0002-8846-9916
Ferdinand Dhombres https://orcid.org/0000-0003-3246-8727

REFERENCES

1. Self A, Daher L, Schlussel M, Roberts N, Ioannou C, Papageorghiou AT. Second and third trimester estimation of gestational age using ultrasound or maternal symphysis-fundal height measurements: a systematic review. BJOG. 2022;129(9):1447–58.
2. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, et al. 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. Lancet. 2012;379(9832):2162–72.
3. Statistics South Africa. Maternal Health Indicators Media Release 2020. Available from: http://www.statssa.gov.za/?p=13100&gclid=CjwKCAjwieuGBhAsEiwA1Ly_nS-sNpoXDRQ9ZN3Z8Cfx8-SoBAJhvXD4wgTWMSagA7hFK4dy-LhHzhoC8X8QAvD_BwE
4. Kaswa R, Rupesinghe GFD, Longo-Mbenza B. Exploring the pregnant women's perspective of late booking of antenatal care services at Mbekweni health Centre in Eastern Cape, South Africa. Afr J Prim Health Care Fam Med. 2018;10(1):e1–e9.
5. Uganda Bureau of Statistics (UBOS) and ICF. Uganda demographic and health survey 2016. Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF; 2018. p. 150.
6. Osterman MJK, Martin JA. Timing and adequacy of prenatal care in the United States, 2016. Natl Vital Stat Rep. 2018;67(3):1–14.
7. NHS Digital. Hospital episode statistics provider level analysis 2017-18 to 2018-19. Leeds: Health and Social Care Information Centre; 2019.
8. Alderliesten ME, Vrijkotte TG, van der Wal MF, Bonsel GJ. Late start of antenatal care among ethnic minorities in a large cohort of pregnant women. BJOG. 2007;114(10):1232–9.
9. Nguyen TH, Larsen T, Engholm G, Moller H. Increased adverse pregnancy outcomes with unreliable last menstruation. Obstet Gynecol. 2000;95(6 Pt 1):867–73.
10. World Health Organisation. WHO recommendations on antenatal care for a positive pregnancy experience 2016. Available from: https://www.who.int/publications/i/item/9789241549912
11. Papageorghiou AT, Kemp B, Stones W, Ohuma EO, Kennedy SH, Purwar M, et al. Ultrasound-based gestational-age estimation in late pregnancy. Ultrasound Obstet Gynecol. 2016;48(6):719–26.
12. Butt K, Lim KI. Guideline No. 388-determination of gestational age by ultrasound. J Obstet Gynaecol Can. 2019;41(10):1497–507.
13. Hadlock FP. Sonographic estimation of fetal age and weight. Radiol Clin North Am. 1990;28(1):39–50.
14. Degani S. Fetal biometry: clinical, pathological, and technical considerations. Obstet Gynecol Surv. 2001;56(3):159–67.
15. Kuleva M, Castaing O, Fries N, Bernard JP, Bussières L, Fontanges M, et al. A standardized approach for the assessment of the lower uterine segment at first trimester by transvaginal ultrasound: a flash study. J Matern Fetal Neonatal Med. 2016;29(9):1376–81.
16. Kuleva M, Fries N, Castaing O, Moeglin D, Salomon LJ. Flash study' on chorionicity determination from ultrasound images at 11-14 weeks' gestation in twin pregnancies. Ultrasound Obstet Gynecol. 2013;41(4):471–2.

17. Stirnemann JJ, Fries N, Bessis R, Fontanges M, Mangione R, Salomon LJ. Implementing the INTERGROWTH-21(st) fetal growth standards in France: a 'flash study' of the college Français d'Echographie Foetale (CFEF). Ultrasound Obstet Gynecol. 2017;49(4):487–92.
18. Salomon LJ. How to date pregnancy? J Gynecol Obstet Biol Reprod. 2011;40(8):726–33.
19. Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. Br J Obstet Gynaecol. 1975;82(9):702–10.
20. Akhtar W, Ali A, Arain MA, Saeed F, Siddiqui S, Memon A. Sonographic fetal biometry charts for a Pakistani cohort. East Mediterr Health J. 2011;17(12):969–75.
21. Akram M, Farooqi U, Arif N, Riaz A, Yousaf M, Fatima M, et al. Estimation of gestational age from fetal kidney length in the second and third trimester of pregnancy by ultrasonography. Saudi Pharm J. 2019;5:222–9.
22. Chervenak FA, Skupski DW, Romero R, Myers MK, Smith-Levitin M, Rosenwaks Z, et al. How accurate is fetal biometry in the assessment of fetal age? Am J Obstet Gynecol. 1998;178(4):678–87.
23. Gupta D, Gupta H, Zaidi Z, Saxena D, Gupta R. Accuracy in estimation of gestational age in third trimester by fetal kidney length in Indian women. Indian J Clin Pract. 2013;24(5):459–63.
24. Hill LM, Guzick D, Hixson J, Peterson CS, Rivello DM. Composite assessment of gestational age: a comparison of institutionally derived and published regression equations. Am J Obstet Gynecol. 1992;166(2):551–5.
25. Kaul I, Menia V, Anand AK, Gupta R. Role of fetal kidney length in estimation of gestational age. JK Sci. 2011;14(2):65–9.
26. Mahale NA, Devi A, Mahale A, Ullal S, Fernandes M, Prabhu S. Ultrasonic assessment of gestational age by measurement of fetal kidney length in third trimester. G Ital Ostet Ginecol. 2019;41(1):10–3.
27. Osho ES, Ibitoye BO, Adetiloye VA, Adeyemi AB, Aderibigbe AS, Omisore AD. Ultrasonic determination of gestational age by assessment of fetal kidney size in the third trimester in southwest Nigeria. Int J Gynaecol Obstet. 2019;144(3):271–6.
28. Ozat M, Kanat-Pektas M, Gungor T, Gurlek B, Caglar M. The significance of fetal sacral length in the ultrasonographic assessment of gestational age. Arch Gynecol Obstet. 2011;283(5):999–1004.
29. Piantelli G, Sacchini C, Coltri A, Ludovici G, Paita Y, Gramellini D. Ultrasound dating-curve analysis in the assessment of gestational age. Clin Exp Obstet Gynecol. 1994;21(2):108–18.
30. Schmidt W, Hendrik H. Fetal femur length in the 2d and 3d trimester of pregnancy. Geburtshilfe Frauenheilkd. 1985;45(2):91–7.
31. Varol F, Saltik A, Kaplan PB, Kilic T, Yardim T. Evaluation of gestational age based on ultrasound fetal growth measurements. Yonsei Med J. 2001;42(3):299–303.
32. Al Hindawi ZM. Ultrasonographic assessment of duration of pregnancy by utilizing fetus kidney length in second and third trimester. Int J Pharm Res. 2020;12(4):1850–6.
33. Altman DG, Chitty LS. New charts for ultrasound dating of pregnancy. Ultrasound Obstet Gynecol. 1997;10(3):174–91.
34. Chang CH, Chang FM, Yu CH, Ko HC, Chen HY. Three-dimensional ultrasound in the assessment of fetal cerebellar transverse and antero-posterior diameters. Ultrasound Med Biol. 2000;26(2):175–82.
35. Chang CH, Chang FM, Yu CH, Yao BL, Ko HC. Fetal head circumference in normal pregnancy: remodeling by Altman's method. J Med Ultrasound. 1998;6(2):61–7.
36. Chang CH, Chang FM, Yu CH, Yao BL, Liang RI, Ko HC. Re-analysis of fetal biparietal diameter during gestation in Taiwanese by Altman's method. J Med Ultrasound. 1996;4(4):162–8.
37. Chang CH, Chang FM, Yu CH, Yao BL, Liang RI, Ko HC, et al. Fetal occipito-frontal diameter in normal pregnancy: reappraisal by Altman's model. J Med Ultrasound. 1997;5(1):5–11.
38. Chang FM, Ko HC, Chang CH, Yu CH, Liang RI, Yao BL, et al. Fetal tibia length assessment by prenatal ultrasound: part I. Using absolute residuals for constructing age-related reference percentiles. J Med Ultrasound. 1997;5(2):49–56.
39. Chavez MR, Ananth CV, Smulian JC, Yeo L, Oyelese Y, Vintzileos AM. Fetal transcerebellar diameter measurement with particular emphasis in the third trimester: a reliable predictor of gestational age. Am J Obstet Gynecol. 2004;191(3):979–84.
40. Dare FO, Smith NC, Smith P. Ultrasonic measurement of biparietal diameter and femur in foetal age determination. West Afr J Med. 2004;23(1):24–6.
41. De Crespigney LC, Speirs AL. A new look at biparietal diameter. Aust N Z J Obstet Gynaecol. 1989;29(1):26–9.
42. Dilmen G, Turhan NO, Toppare MF, Seckin N, Ozturk M, Goksin E. Scapula length measurement for assessment of fetal growth and development. Ultrasound Med Biol. 1995;21(2):139–42.
43. Edevbie JP, Akhigbe AO. Ultrasound measurement of fetal kidney length in normal pregnancy and correlation with gestational age. Niger J Clin Pract. 2018;21(8):960–6.
44. Eze CU, Onu IU, Adeyomoye AA, Upeh ER. Estimation of gestational age using trans-cerebellar diameter: a sonographic study of a cohort of healthy pregnant women of Igbo ethnic origin in a suburb of Lagos, southwest Nigeria. J Ultrasound. 2020;19:19.
45. Eze CU, Onwuzu QE, Nwadike IU. Sonographic reference values for fetal transverse cerebellar diameter in the second and third trimesters in a Nigerian population. J Diagn Med Sonogr. 2017;33(3):174–81.
46. Gameraddin G, Suliman S, Yousef Y. Evaluation of gestational age with fetal foot length using ultrasonography. J Am Sci. 2014;10:5–7.
47. Ghai R, Singla M, Goel P, Khare S, Jain S, Agrawal S. Ultrasonic measurement of foetal biparietal diameter and its correlation to gestational age in the garhwali population. Indian J Public Health Res Dev. 2011;2(1):74–7.
48. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal biparietal diameter: a critical re-evaluation of the relation to menstrual age by means of real-time ultrasound. J Ultrasound Med. 1982;1(3):97–104.
49. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal abdominal circumference as a predictor of menstrual age. AJR Am J Roentgenol. 1982;139(2):367–70.
50. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal head circumference: relation to menstrual age. AJR Am J Roentgenol. 1982;138(4):649–53.
51. Hadlock FP, Deter RL, Harrist RB, Park SK. Computer assisted analysis of fetal age in the third trimester using multiple fetal growth parameters. J Clin Ultrasound. 1983;11(6):313–6.
52. Hadlock FP, Harrist RB, Deter RL, Park SK. Fetal femur length as a predictor of menstrual age: sonographically measured. AJR Am J Roentgenol. 1982;138(5):875–8.
53. 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.
54. Hegab M, Midan MF, Taha T, Bibars M, Wakeel KHE, Amer H, et al. Fetal biometric charts and reference equations for pregnant women living in Port Said and Ismailia governorates in Egypt. Open Access Maced J Med Sci. 2018;6(5):751–6.
55. Hemraj S, Abraham SM, Acharya D, Ravichandra G, Vinayaka US. Sonographic correlation of gestational age with fetal kidney length. Int J Infertil Fetal Med. 2016;7(2):37–41.
56. Holanda-Filho JA, Souza AI, Souza AS, Figueroa JN, Ferreira AL, Cabral-Filho JE. Fetal transverse cerebellar diameter measured by ultrasound does not differ between genders. Arch Gynecol Obstet. 2011;284(2):299–302.
57. Honarvar M, Allahyari M, Dehbashi S. Assessment of gestational age based on ultrasonic femur length after the first trimester: a simple mathematical correlation between gestational age (GA) and femur length (FL). Int J Gynaecol Obstet. 2000;70(3):335–40.
58. Jiang X, Zhang YH, Li Y, Ma X, Zhu YS, Shang L. Reference charts and equations of fetal biometry for normal singleton

pregnant women in Shaanxi, China. Clin Exp Obstet Gynecol. 2013;40(3):393–8.
59. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on ultrasound head biometry and the effect of maternal and fetal factors. Acta Obstet Gynecol Scand. 2004;83(8):716–23.
60. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on femur length at 10-25 weeks of gestation, and reference ranges for femur length to head circumference ratios. Acta Obstet Gynecol Scand. 2005;84(8):725–33.
61. Karabulut AK, Koyluoglu B, Uysal I. Human foetal sacral length measurement for the assessment of foetal growth and development by ultrasonography and dissection. Anat Histol Embryol. 2001;30(3):141–6.
62. Konje JC, Abrams KR, Bell SC, Taylor DJ. Determination of gestational age after the 24th week of gestation from fetal kidney length measurements. Ultrasound Obstet Gynecol. 2002;19(6):592–7.
63. Leung TN, Pang MW, Daljit SS, Leung TY, Poon CF, Wong SM, et al. Fetal biometry in ethnic Chinese: biparietal diameter, head circumference, abdominal circumference and femur length. Ultrasound Obstet Gynecol. 2008;31(3):321–7.
64. Mahmoud MZ, Mahmoud O, Abdulla A. Fetal transverse cerebellar diameter measurement for prediction of gestational age in pregnant Sudanese ladies. Int J Life Sci Med Res. 2013;3:89–93.
65. Marinho AO, Bamgboye EA. Assessment of foetal femur length by ultrasound in a normal Nigerian obstetric population. Afr J Med Med Sci. 1987;16(2):47–52.
66. AMANHI. Performance of late pregnancy biometry for gestational age dating in low-income and middle-income countries: a prospective, multicountry, population-based cohort study from the WHO Alliance for maternal and newborn health improvement (AMANHI) study group. Lancet Global Health. 2020;8(4):e545–e554.
67. Mercer BM, Sklar S, Shariatmadar A, Gillieson MS, D'Alton ME. Fetal foot length as a predictor of gestational age. Am J Obstet Gynecol. 1987;156(2):350–5.
68. Mongelli M, Chew S, Yuxin NG, Biswas A. Third-trimester ultrasound dating algorithms derived from pregnancies conceived with artificial reproductive techniques. Ultrasound Obstet Gynecol. 2005;26(2):129–31.
69. Nwadike UI, Agwu KK, Eze CU, Okpala OC, Onu AO. Sonographic biometry of Fetal interorbital distance as a predictor of gestational age in Enugu, Southeast Nigeria. West Indian Med J. 2016;65(1):128–33.
70. Orji MO, Adeyekun AA. Ultrasound estimation of foetal gestational age by transcerebellar diameter in healthy pregnant nigerian women. West Afr J Med. 2014;33(1):61–7.
71. Osinusi BO, Ogunseyinde O. Ultrasonic foetal abdominal circumference as a means of assessing gestational age in Nigerians. Afr J Med Med Sci. 1989;18(2):101–4.
72. Osinusi BO, Okubanjo OA. Ultrasonic foetal head circumference as a means of assessing gestational age in Nigerians. West Afr J Med. 1990;9(1):22–5.
73. Ott WJ. Accurate gestational dating: revisited. Am J Perinatol. 1994;11(6):404–8.
74. Ott WJ. Accurate gestational dating. Obstet Gynecol. 1985;66(3):311–5.
75. Persson PH, Weldner BM. Reliability of ultrasound fetometry in estimating gestational age in the second trimester. Acta Obstet Gynecol Scand. 1986;65(5):481–3.
76. Persson PH, Weldner BM. Normal range growth curves for fetal biparietal diameter, occipito frontal diameter, mean abdominal diameters and femur length. Acta Obstet Gynecol Scand. 1986;65(7):759–61.
77. Quinlan RW, Brumfield C, Martin M, Cruz AC. Ultrasonic measurement of femur length as a predictor of fetal gestational age. J Reprod Med. 1982;27(7):392–4.
78. Reddy RH, Prashanth K, Ajit M. Significance of foetal transcerebellar diameter in foetal biometry: a pilot study. J Clin Diagn Res. 2017;11(6):TC01–C04.
79. Rodriguez-Sibaja MJ, Villar J, Ohuma EO, Napolitano R, Heyl S, Carvalho M, et al. Fetal cerebellar growth and Sylvian fissure maturation: international standards from fetal growth longitudinal study of INTERGROWTH-21st project. Ultrasound Obstet Gynecol. 2021;57(4):614–23.
80. Sabbagha RE, Turner JH, Rockette H, Mazer J, Orgill J. Sonar BPD and fetal age. Definition of the relationship. Obstet Gynecol. 1974;43(1):7–14.
81. Satish Prasad BS, Likhitha S. Cerebellar measurements with ultrasonography in the evaluation of fetal age. IOSR J Dent Med Sci. 2014;13(9):49–56.
82. Schluter PJ, Pritchard G, Gill MA. Using ultrasonic fetal size measurements to estimate gestational age in Brisbane, Australia. Australas Radiol. 2007;51(1):46–52.
83. Selbing A, Kjessler B. Conceptual dating by ultrasonic measurement of the fetal biparietal diameter in early pregnancy. Acta Obstet Gynecol Scand. 1985;64(7):593–7.
84. Senapati S, Behera SS, Chinara PK. Relationship of umbilical cord size with gestational age: a sonographic study. Paripex Indian J Res. 2016;5(4):434–9.
85. Skupski DW, Owen J, Kim S, Fuchs KM, Albert PS, Grantz KL, et al. Estimating gestational age from ultrasound fetal biometrics. Obstet Gynecol. 2017;130(2):433–41.
86. Sun JF, Zhang LJ, Cui LQ, Luo XD, He GH, Dong XJ, et al. Gestational age estimation from ultrasound fetal biometrics in China. Clin Exp Obstet Gynecol. 2020;47(5):714–22.
87. Swapnil P, Vilas C, Varsha P. Correlation between gestational age and abdominal circumference in second trimester. Indian J Forensic Med Toxicol. 2021;15(1):440–4.
88. Tongsong T, Wanapirak C, Jesadapornchai S, Tathayathikom E. Fetal binocular distance as a predictor of menstrual age. Int J Gynaecol Obstet. 1992;38(2):87–91.
89. Tse CH, Lee KW. A comparison of the fetal femur length and biparietal diameter in predicting gestational age in the third trimester. Aust N Z J Obstet Gynaecol. 1984;24(3):186–8.
90. Uerpairojkit B, Charoenvidhya D, Manotaya S, Tanawattanachareon S, Wacharaprechanont T, Tannirandorn Y. Fetal transverse cerebellar diameter in Thai population. J Med Assoc Thai. 2001;84(Suppl 1):S346–S351.
91. Ugur MG, Mustafa A, Ozcan HC, Tepe NB, Kurt H, Akcil E, et al. Fetal kidney length as a useful adjunct parameter for better determination of gestational age. Saudi Med J. 2016;37(5):533–7.
92. Varsha P, Swapnil P, Bharat P. Correlation between gestational age and head circumference in third trimester. Indian J Forensic Med Toxicol. 2020;14(4):273–9.
93. Verburg BO, Steegers EA, De Ridder M, Snijders RJ, Smith E, Hofman A, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obstet Gynecol. 2008;31(4):388–96.
94. Viveki PR, Shirol VS. Establishment of reference tables for fetal abdominal circumference by ultrasonographic measurements in local population, Karnataka, India. Indian J Public Health Res Dev. 2019;10(6):44–50.
95. Warda AH, Deter RL, Rossavik IK, Carpenter RJ, Hadlock FP. Fetal femur length: a critical reevaluation of the relationship to menstrual age. Obstet Gynecol. 1985;66(1):69–75.
96. Westerway SC, Davison A, Cowell S. Ultrasonic fetal measurements: new Australian standards for the new millennium. Aust N Z J Obstet Gynaecol. 2000;40(3):297–302.
97. Yagel S, Adoni A, Oman S, Wax Y, Hochner-Celnikier D. A statistical examination of the accuracy of combining femoral length and biparietal diameter as an index of fetal gestational age. Br J Obstet Gynaecol. 1986;93(2):109–15.

98. 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):133–46.
99. Chavez MR, Ananth CV, Smulian JC, Vintzileos AM. Fetal transcerebellar diameter measurement for prediction of gestational age at the extremes of fetal growth. J Ultrasound Med. 2007;26(9):1167; quiz 73–4–71.
100. McMillen IC, Adams MB, Ross JT, Coulter CL, Simonetta G, Owens JA, et al. Fetal growth restriction: adaptations and consequences. Reproduction. 2001;122(2):195–204.
101. Koech A, Musitia PM, Mwashigadi GM, Kinshella MW, Vidler M, Temmerman M, et al. Acceptability and feasibility of a low-cost device for gestational age assessment in a low-resource setting: qualitative study. JMIR Hum Factors. 2022;9(4):e34823.
102. Chamberlain PF, Manning FA, Morrison I, Harman CR, Lange IR. Ultrasound evaluation of amniotic fluid volume. I. The relationship of marginal and decreased amniotic fluid volumes to perinatal outcome. Am J Obstet Gynecol. 1984;150(3):245–9.
103. Zilianti M, Fernández S. Correlation of ultrasonic images of fetal intestine with gestational age and fetal maturity. Obstet Gynecol. 1983;62(5):569–73.
104. Gentili P, Trasimeni A, Giorlandino C. Fetal ossification centers as predictors of gestational age in normal and abnormal pregnancies. J Ultrasound Med. 1984;3(5):193–7.
105. Petrucha RA, Platt LD. Relationship of placental grade to gestational age. Am J Obstet Gynecol. 1982;144(6):733–5.
106. Malik R, Pandya VK, Shrivastava P. Gestational age estimation using transcerebellar diameter with grading of fetal cerebellum and evaluation of TCD/AC (Transcerebellar diameter/abdominal circumference) ratio as a gestational age independent parameter. Indian J Radiol Imag. 2003;13(1):95–7.
107. Hashimoto K, Shimizu T, Shimoya K, Kanzaki T, Clapp JF, Murata Y. Fetal cerebellum: US appearance with advancing gestational age. Radiology. 2001;221(1):70–4.
108. Lee LH, Bradburn E, Craik R, Yaqub M, Norris SA, Ismail LC, et al. Machine learning for accurate estimation of fetal gestational age based on ultrasound images. NPJ Digit Med. 2023;6(1):36.
109. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Vellvé K, Eixarch E, Crispi F, et al. Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. Am J Obstet Gynecol. 2021;3(6):100462.
110. Fung R, Villar J, Dashti A, Ismail LC, Staines-Urias E, Ohuma EO, et al. Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. Lancet Digital Health. 2020;2(7):e368–e375.
111. Napolitano R, Donadono V, Ohuma EO, Knight CL, Wanyonyi SZ, Kemp B, et al. Scientific basis for standardization of fetal head measurements by ultrasound: a reproducibility study. Ultrasound Obstet Gynecol. 2016;48(1):80–5.

Napolitano et al.[111] cited in Files S5 and S6.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Self A, Schlussel M, Collins GS, Dhombres F, Fries N, Haddad G, et al. External validation of models to estimate gestational age in the second and third trimester using ultrasound: A prospective multicentre observational study. BJOG. 2024;00:1–12. https://doi.org/10.1111/1471-0528.17922