International estimated fetal weight standards of the INTERGROWTH-21st Project

J. STIRNEMANN[1], J. VILLAR[2]#, L. J. SALOMON[1,3], E. OHUMA[2,4], P. RUYAN[5], D. G. ALTMAN[4], F. NOSTEN[6], R. CRAIK[2], S. MUNIM[7], L. CHEIKH ISMAIL[2], F. C. BARROS[8,9], A. LAMBERT[2], S. NORRIS[10], M. CARVALHO[11], Y. A. JAFFER[12], J. A. NOBLE[13], E. BERTINO[14], M. G. GRAVETT[15], M. PURWAR[16], C. G. VICTORA[9], R. UAUY[17,18]#, Z. BHUTTA[19]#, S. KENNEDY[2]# and A. T. PAPAGEORGHIOU[2]#, for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st)

[1]Maternité Necker-Enfants Malades, AP-HP & EA7328 Université Paris Descartes, Paris, France; [2]Nuffield Department of Obstetrics & Gynaecology and Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK; [3]Collège Français d'Echographie Foetale – CFEF, France; [4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK; [5]School of Public Health, Peking University, Beijing, China; [6]Shoklo Malaria Research Unit, Maesod, Tak, Thailand; [7]Division of Women & Child Health, The Aga Khan University, Karachi, Pakistan; [8]Programa de Pós-Graduação em Saúde e Comportamento, Universidade Católica de Pelotas, Pelotas, RS, Brazil; [9]Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil; [10]Developmental Pathways For Health Research Unit, Department of Paediatrics & Child Health, University of the Witwatersrand, Johannesburg, South Africa; [11]Faculty of Health Sciences, Aga Khan University, Nairobi, Kenya; [12]Department of Family & Community Health, Ministry of Health, Muscat, Sultanate of Oman; [13]Department of Engineering Science, University of Oxford, Oxford, UK; [14]Dipartimento di Scienze Pediatriche e dell'Adolescenza, Cattedra di Neonatologia, Università degli Studi di Torino, Torino, Italy; [15]Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), Seattle, WA, USA; [16]Nagpur INTERGROWTH-21st Research Centre, Ketkar Hospital, Nagpur, India; [17]Division of Paediatrics, Pontifical Universidad Catolica de Chile, Chile; [18]London School of Hygiene and Tropical Medicine, London, UK; [19]Center for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada

ABSTRACT

Objective Estimated fetal weight (EFW) and fetal biometry are complementary measures used to screen for fetal growth disturbances. Our aim was to provide international EFW standards to complement the INTERGROWTH-21st Fetal Growth Standards that are available for use worldwide.

Methods Women with an accurate gestational-age assessment, who were enrolled in the prospective, international, multicenter, population-based Fetal Growth Longitudinal Study (FGLS) and INTERBIO-21st Fetal Study (FS), two components of the INTERGROWTH-21st Project, had ultrasound scans every 5 weeks from 9–14 weeks' until 40 weeks' gestation. At each visit, measurements of fetal head circumference (HC), biparietal diameter, occipitofrontal diameter, abdominal circumference (AC) and femur length (FL) were obtained blindly by dedicated research sonographers using standardized methods and identical ultrasound machines. Birth weight was measured within 12 h of delivery by dedicated research anthropometrists using standardized methods and identical electronic scales. Live babies without any congenital abnormality, who were born within 14 days of the last ultrasound scan, were selected for inclusion. As most births occurred at around 40 weeks' gestation, we constructed a bootstrap model selection and estimation procedure based on resampling of the complete dataset under an approximately uniform distribution of birth weight, thus enriching the sample size at extremes of fetal sizes, to achieve consistent estimates across the full range of fetal weight. We constructed reference centiles using second-degree fractional polynomial models.

Results Of the overall population, 2404 babies were born within 14 days of the last ultrasound scan. Mean time between the last scan and birth was 7.7 (range, 0–14) days and was uniformly distributed. Birth weight was best estimated as a function of AC and HC (without FL) as $\log(EFW) = 5.084820 - 54.06633 \times$

Correspondence to: Prof. A. T. Papageorghiou, Nuffield Department of Obstetrics and 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)

#J.V., R.U., Z.B., S.K. and A.T.P. are joint senior authors.

ORIGINAL PAPER

$(AC/100)^3 - 95.80076 \times (AC/100)^3 \times log(AC/100) + 3.136370 \times (HC/100)$, *where EFW is in g and AC and HC are in cm. All other measures, gestational age, symphysis–fundus height, amniotic fluid indices and interactions between biometric measures and gestational age, were not retained in the selection process because they did not improve the prediction of EFW. Applying the formula to FGLS biometric data (n = 4231) enabled gestational age-specific EFW tables to be constructed. At term, the EFW centiles matched those of the INTERGROWTH-21st Newborn Size Standards but, at < 37 weeks' gestation, the EFW centiles were, as expected, higher than those of babies born preterm. Comparing EFW cross-sectional values with the INTERGROWTH-21st Preterm Postnatal Growth Standards confirmed that preterm postnatal growth is a different biological process from intrauterine growth.*

Conclusions We provide an assessment of EFW, as an adjunct to routine ultrasound biometry, from 22 to 40 weeks' gestation. However, we strongly encourage clinicians to evaluate fetal growth using separate biometric measures such as HC and AC, as well as EFW, to avoid the minimalist approach of focusing on a single value. © 2016 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

INTRODUCTION

One of the main objectives of antenatal care is screening for fetal growth disturbances[1]. Although biomarkers in maternal blood have shown some potential for detecting fetal growth restriction[2,3], a recent systematic review suggested that none is sufficiently accurate to be recommended for use in clinical practice[4]. Clinicians, therefore, still rely routinely on clinical markers, including ultrasound measurements, to identify fetuses at risk[5].

Ultrasound evaluation of the fetus involves measuring head circumference (HC), abdominal circumference (AC) and femur length (FL), and the values can be combined to calculate an estimated fetal weight (EFW); this estimate is often used alone in clinical practice without considering the individual measurements. However, we believe that arguments concerning the most appropriate single parameter to use are inappropriate because clinicians should use all the tools available in their armamentarium for making crucial clinical decisions that have major implications for both mothers and newborns.

The development of international EFW standards is overdue, and these should share the same conceptual basis as the published INTERGROWTH-21st standards for HC, AC and FL, size at birth and postnatal growth in preterm infants[6–9]. These standards would perfectly complement the World Health Organization (WHO) Child Growth Standards[10], thereby enabling continuity of assessment of human growth from early pregnancy to childhood[11]. Therefore, the objectives of this component of the INTERGROWTH-21st Project were: (1) to develop a formula to estimate fetal weight based on ultrasound biometry and birth weight; and (2) to construct international EFW standards for fetuses at 22 to 40 weeks' gestation.

SUBJECTS AND METHODS

INTERGROWTH-21st is an international, multicenter, population-based project consisting of a number of components, including the Fetal Growth Longitudinal Study (FGLS) and INTERBIO-21st Fetal Study (FS).

FGLS was conducted between 27 April 2009 and 2 March 2014 in eight urban areas: the cities of Pelotas (Brazil), Turin (Italy), Muscat (Oman), Oxford (UK) and Seattle (USA); the Shunyi County of Beijing (China); the central area of Nagpur (India); and the Parklands suburb of Nairobi (Kenya). The primary aim was to study longitudinally the health and development of fetuses into infancy, by monitoring growth, health, nutrition and neurodevelopment from less than 14 weeks' gestation to 2 years of age, so as to produce prescriptive growth standards to complement the existing WHO Child Growth Standards. This was achieved by studying a cohort of healthy, well-nourished, pregnant women who were at low risk of adverse maternal and perinatal outcomes at both population and individual levels. The study details have been described elsewhere[9,12].

In contrast, FS recruited an unselected cohort of pregnant women, between 8 February 2012 and 24 December 2015, from three FGLS sites (Pelotas, Nairobi, Oxford), and three new sites (Aga Khan University Hospital, Karachi, Pakistan; Shoklo Malaria Research Unit, Mae Sot, Thailand; and Baragwanath Hospital, Soweto, South Africa). The primary aim was to study the effects of various intrauterine exposures (e.g. malnutrition, anemia, human immunodeficiency virus, malaria) on growth, health, nutrition, neurodevelopment and the epigenome, over the same developmental age range, i.e. from less than 14 weeks' gestation to 2 years of age.

To develop the EFW formula requires as many pregnancies as possible that have a standardized scan and birth-weight measurement. In order to achieve this we included fetuses from both FGLS and FS; only those that had an ultrasound scan within 14 days of birth were included in the calculations. To develop the international standards for EFW, the formula derived was then applied to the healthy FGLS population from which the International Fetal Growth Standards were produced[6].

The INTERGROWTH-21st Project was approved by the Oxfordshire Research Ethics Committee 'C' (reference: 08/H0606/139), the research ethics committees of the individual participating institutions and the corresponding regional health authorities in which the project was implemented. Participants provided written consent to be involved in the study.

Standard procedures

In both studies women were recruited at less than 14 weeks' gestation. All women underwent ultrasound measurement of fetal crown–rump length (CRL) using standardized methodology[13,14]. In FGLS, gestational age was based on the date of the last menstrual period (LMP) provided it was certain, the woman had a regular 24–32-day menstrual cycle and she had not been using hormonal contraception or breastfeeding in the preceding 2 months, and any discrepancy between the gestational ages based on LMP and CRL, between 9 + 0 and 13 + 6 weeks, was ≤ 7 days. In FS, gestational age was determined by CRL measurement alone, using the same formula loaded onto all study ultrasound machines[15]; if known, the date of the LMP was recorded.

Following the dating scan, women were scanned every 5 weeks (±1 week), so that the possible ranges were 14–18, 19–23, 24–28, 29–33, 34–38 and 39–42 weeks' gestation. At each visit, fetal HC, biparietal diameter (BPD), occipitofrontal diameter (OFD), AC and FL were measured three times from three separately obtained ultrasound images of each structure. The detailed measurement protocols, including graphical displays of measurement techniques, and the unique standardization procedures for all measurements and sonographer training have been reported elsewhere[13,16]. In addition, all documentation, protocols, quality-control procedures, data collection forms and electronic transfer strategies are freely available on the INTERGROWTH-21st website.

Briefly, 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 screen. The midline echo (representing the falx cerebri) had to be broken anteriorly, at a third of its length, by the cavum septi pellucidi. The thalami had to be located symmetrically on either side of the midline. Calipers 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; HC was measured using the ellipse facility on the outer border of the skull.

AC measurements were taken in a cross-sectional view of the fetal abdomen as close as possible to circular in shape, with the umbilical vein in the anterior third (at the level of the portal sinus), with the stomach bubble visible. The sonographer was instructed to avoid applying too much pressure with the transducer, which can distort the circular shape of the fetal abdomen. The abdomen had to fill at least 30% of the monitor screen, and the spine had to be at either the 3 or 9 o'clock position 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, 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 about 90°, with the full length of the bone visualized, unobscured by shadowing from adjacent bony parts, and the femur had to fill at least 30% of the monitor screen. The intersection of the calipers was placed on the outer borders of the edges of the femoral diaphysis (outer to outer) ensuring clear femoral edges; ultrasound artifacts of the femoral edges such as the proximal trochanter or pointed femoral spurs were not included in the measurement (detailed methods and a graphical display of how the bone structures are localized are available on the INTERGROWTH-21st website).

The same type of ultrasound machine, a Philips HD-9 with curvilinear abdominal transducers C5-2, C6-3 and V7-3 (Philips Ultrasound, Bothell, WA, USA), was used at all sites. To avoid expected value bias, the machine was adapted so that fetal measurements were not visible to the sonographer on the screen. Only after three measurements of each structure had been recorded were the average values revealed for clinical purposes. All ultrasound data were submitted electronically to the study database. Data were entered locally directly onto the web-based system[17].

After taking each set of measurements, sonographers scored the quality of their images on the basis of standard image-scoring criteria[18,19]. Images that did not score the maximum number of points were repeated until the best possible score was achieved. The quality-control methods used across all sites are described in detail elsewhere[18,20].

Birth weight was measured within 12 h of birth using identical electronic scales (Seca, Hangzhou, China) at all sites. The equipment, which was calibrated twice a week, was selected for accuracy, precision and robustness, as shown previously[21]. Measurement procedures were standardized on the basis of WHO recommendations to ensure maximum validity and each measurement was collected independently by two study anthropometrists[22,23]. If the difference between the two measurements exceeded the maximum allowable difference of 5 g, then both observers independently retook that measurement a second and, if necessary, a third time. The training, standardization, monitoring processes and quality-control methods used across all sites are described in detail elsewhere[22,23].

Statistical analysis

Estimation of fetal weight

From the FGLS and FS cohorts, we identified all live babies without any congenital abnormality who were born at > 24 weeks' gestation and within 14 days of the last ultrasound scan. Given the study design, we expected the births to have occurred uniformly between 0 and 14 days after the last ultrasound scan, i.e. we expected there to be a mean time of 7–8 days between the last scan and birth. This cut-off allowed a greater number of births at low gestational ages to be included, for which most of the existing formulae have been prone to prediction error, probably because scant data exist for estimation[24]. Potential predictors for birth weight were:

- HC, BPD, OFD, AC and FL, in cm or transformed into Z-scores using the INTERGROWTH-21st equations[6];
- gestational age on the day of the last scan, in weeks;
- symphysis–fundus height, in cm;
- amniotic fluid, assessed by the deepest vertical pool and amniotic fluid index in cm;
- cross-sectional head area and abdominal area computed from their orthogonal diameters, in cm[2].

We hypothesized that the contribution of any given anthropometric measurement to EFW might vary with gestational age. Therefore, we also considered interactions between HC, BPD, OFD, AC and FL and gestational age on the day of the last scan. Statistical modeling was conducted using second-degree fractional polynomials[25].

Some prediction bias would be expected because of significant growth between the day of the last scan and birth[24,26,27]. We addressed this issue by calculating the expected EFW on the day of the ultrasound scan, using the following steps: (1) in pregnancies from FGLS and FS delivering within 14 days from the last scan, we developed a model to predict birth weight from the most recent ultrasound measurements; (2) in the complete FGLS dataset, we calculated EFW from ultrasound biometry using the previous model and fitted a second-degree fractional polynomial for mean weight as a function of gestational age between 22 and 40 weeks; (3) returning to the dataset of births within 14 days (step 1), we calculated, for each fetus, the expected weight at the time of the last scan by subtracting the average weight gain between the time of the last scan and birth using the model built in step 2; (4) this calculated weight was then used for further modeling.

As expected, owing to the prospective, population-based design of FGLS, most births occurred close to 40 weeks' gestation, meaning that the scatter of observations across the 22–40-week window was very uneven. We were aware that estimation using the complete dataset would yield very accurate estimates at 40 weeks' gestation, where the greatest contribution of the data is found, but with limited model validity for lower birth weights. To overcome this problem and allow accurate birth-weight estimation over the whole range of observed data, we constructed a bootstrap model selection and estimation procedure based on resampling of the complete dataset under an approximately uniform distribution of birth weight[28–30], i.e. birth weight was divided into 500-g strata and each sample was built by randomly selecting five observations with replacement from each stratum. In a first resampling run of 100 samples, candidate models, which include three elements (the variables, the coefficients and the respective fractional polynomial powers), were elicited using the backward elimination algorithm described by Ambler and Royston[31], which provides protection against over-fitting. In a second step, the coefficients of all candidate models were estimated in B = 1000 bootstrap samples: in each sample, a single model was selected using Akaike's Information Criterion.

Candidate models were then ordered by their frequency of selection within the 1000 samples, and the five most frequent models were kept for further assessment of goodness of fit.

Assessment of goodness of fit in the complete dataset relied on inspection of residuals with quantile–quantile (q-q) plots and residuals *vs* fitted plots. Given that we estimated fetal weight at the time of the last scan using an average model for growth, we investigated the bias of our model for EFW by calculating the mean of percent prediction errors defined by the formula (100 × (EFW − birth weight)/birth weight), for decreasing time-to-birth intervals (i.e. from 14 to 0 days). Finally, we also calculated the absolute percent prediction error defined by the mean of the absolute prediction errors.

Construction of reference centiles

The construction of reference centiles was based solely on FGLS data. The sample size was based on pragmatic and statistical considerations; the latter focused on the precision and accuracy of one extreme centile, i.e. the 3rd or 97th centile, and regression-based reference limits[32,33]. We have shown that a sample of 4000 women would obtain a precision of 0.03 SD at the 3rd or 97th centile. Further details on the precision obtained at the 5th or 10th centile by sample size (ranging from 500–6000) have been included in a table in a previous publication[34].

The data from all the study sites were pooled to construct the Fetal Growth Standards[6,12], using the same statistical approach adopted by WHO in constructing their Child Growth Standards[10]. The statistical methods used were based on published recommendations complemented by recent scientific reviews[5,35–38]. Our overall aim was to produce centiles that change smoothly with age and maximize simplicity without compromising model fit.

We explored the following statistical methods: mean and SD method using fractional polynomials[25]; Cole's lambda (λ), mu (μ), and sigma (σ) (LMS) method[39–41], which estimates three age-specific parameters (the median (μ), coefficient of variation (σ), and a Box–Cox power transformation at each gestational age to remove skewness (λ), thereby making the data roughly normally distributed); the LMST method (λ, μ, σ, assuming Box–Cox t distribution), which assumes a shifted and scaled (truncated) t distribution to take account of skewness and leptokurtosis[42]; and the LMSP method (λ, μ, σ, assuming Box–Cox power exponential distribution), which assumes a Box–Cox power exponential distribution to take account of skewness, platykurtosis and leptokurtosis[43]. Furthermore, to present the curves, we assessed three smoothing techniques: fractional polynomials, cubic splines and penalized splines[25,44,45].

Using de-trended q-q plots (worm plots), significant evidence of deviations from normality was seen so we

resorted to using the more complex LMS, LMST and LMSP methods allowing for skewness and kurtosis[46].

As most of the women had four to six ultrasound scans, the effect of correlated data within fetuses was investigated. First, in a sensitivity analysis, a random observation time was sampled for each fetus and the modeled centiles in this subset were compared visually with the complete dataset. The approach is justified by the experimental design of the study, which ensures non-informative observation times[47]. This analysis showed minimal or no change in estimated centiles (median, 3rd and 97th centiles) over the whole 22–40 weeks' gestational-age range. Second, we considered mixed-effect models accounting for repeated measurements within LMS, LMST and LMSP frameworks. This analysis also showed no impact on the estimated centiles. The best fit was found using a three-parameter Box–Cox Gaussian distribution (i.e. the LMS method) for the response variable with a second-degree fractional polynomial functional form for gestational age. This method also gives estimated SDs of EFW, allowing estimation of centiles.

Goodness of fit for the overall model was assessed by comparing empirical centiles (calculated per completed week of gestation) with the fitted centiles, using de-trended q-q plots of the residuals across gestational age[46], and plots of residuals *vs* fitted values.

All analyses were carried out in R statistical software[48] using the Generalised Additive Models for Location, Scale and Shape (GAMLSS) framework[49,50].

RESULTS

To create an EFW formula, the subsets of 2404 babies in the FGLS ($n = 1556$) and FS ($n = 848$) who were born within 14 days of the last ultrasound scan were examined (Table 1): 130 (5.4%) were born preterm (< 37 weeks' gestation) and 78 (3.2%) were born term with low birth weight (< 2500 g and ≥ 37 weeks' gestation). The mean time between the last ultrasound scan and birth was 7.7 (range, 0–14) days and was uniformly distributed, except for 0 days (i.e. birth on day of last scan), which occurred in only 34 (1.4%) cases.

Following correction for potential growth between the last scan and birth (steps 1–4 in the statistical methods), the actual fetal weight at the time of the last scan was best estimated as a function of AC and HC with the following formula:

$$\log(\text{EFW}) = 5.084820 - 54.06633 \times (\text{AC}/100)^3$$
$$- 95.80076 \times (\text{AC}/100)^3 \times \log(\text{AC}/100)$$
$$+ 3.136370 \times (\text{HC}/100)$$

where EFW is expressed in g, AC and HC in cm, and the log function designates the natural logarithm.

None of the other covariates including FL, BPD, OFD, gestational age, symphysis–fundus height, amniotic fluid indices or interactions between biometric measurements and gestational age was retained in the selection process.

Table 1 Gestational age at birth and birth weight of a subset of babies in the INTERBIO-21st Fetal Study (FS) and the Fetal Growth Longitudinal Study (FGLS) of the INTERGROWTH-21st Project who were born within 14 days of last ultrasound scan

Parameter	FS (n = 848)	FGLS (n = 1556)	Total (n = 2404)
Gestational age at birth			
< 28 weeks	2 (0.2)	1 (0.1)	3 (0.1)
28–32 weeks	3 (0.4)	6 (0.4)	9 (0.4)
32–37 weeks	56 (6.6)	62 (4.0)	118 (4.9)
37–43 weeks	787 (92.8)	1487 (95.6)	2274 (94.6)
Birth weight			
< 1000 g	1 (0.1)	2 (0.1)	3 (0.1)
1000–1499 g	5 (0.6)	1 (0.1)	6 (0.2)
1500–1999 g	9 (1.1)	15 (1.0)	24 (1.0)
2000–2499 g	79 (9.3)	76 (4.9)	155 (6.4)
≥ 2500 g	754 (88.9)	1462 (94.0)	2216 (92.2)

Data are given as *n* (%).

This model suggests a linear relationship between log(EFW) and HC. Despite the negative coefficients, the two terms involving AC describe an increasing sigmoid-shaped relationship between AC and birth weight (Figure S1) for a fixed HC value of 26 cm (the average value at 28 weeks' gestation[6]). The relationship between birth weight and HC is plotted in Figure S2, for a fixed AC value of 23 cm (the average at 28 weeks' gestation[6]).

The performance of the formula for EFW was assessed both by mean and absolute percent prediction errors; mean percent prediction error is used as a measure of potential bias of EFW due to growth between the last scan and birth, while mean absolute prediction error represents the dispersion of the errors. The mean percent prediction error steadily tended towards zero as the time interval between the last scan and birth decreased. Prediction error was −10.7% (95% CI, −12.1 to −9.4%) in babies born exactly 14 days after the last scan ($n = 196$) and −0.8% (95% CI, −2.3 to 0.6%) in those born within 1 day ($n = 198$) (Figure S3), showing that our model was unbiased for predicting weight at the time of the last scan and that the correction we applied to compensate for time to birth was appropriate. In the group born within 1 day of the last scan, the mean absolute prediction error was 7.6%, with 80%, 90% and 95% of predicted weights falling within 11%, 14% and 18% of the true birth weight, respectively.

Creation of the international EFW standards was based on the complete FGLS dataset. The gestational age-specific observed and smoothed centiles for EFW are presented in Figure 1. Similarities between smoothed centile curves (3rd, 50th and 97th centiles) and observed values, assessed by gestational age-specific comparisons, demonstrated excellent agreement. The overall differences between empirical and smoothed centiles were small, with mean ± SD differences of 16 ± 28 g, 13 ± 17 g and 5 ± 33 g for the 3rd, 50th and 97th centiles, respectively.

The 3rd, 10th, 50th, 90th and 97th fitted centile curves for EFW according to gestational age, which represent
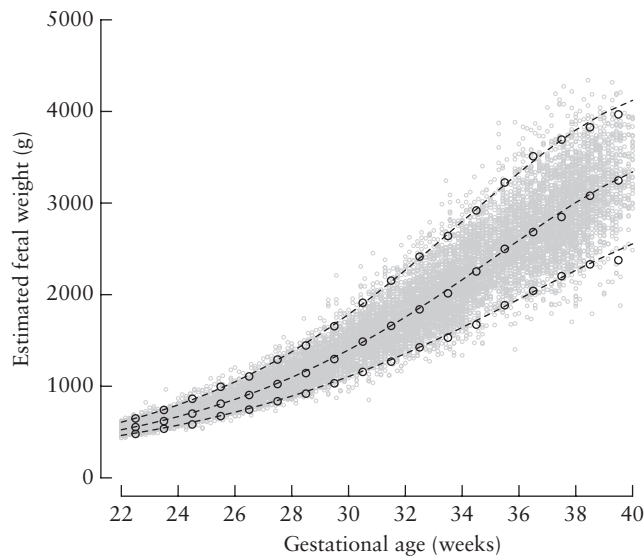
Figure 1 Empirical (○) and smoothed (-----) 3rd, 50th and 97th centiles for estimated fetal weight between 22 and 40 weeks' gestation.
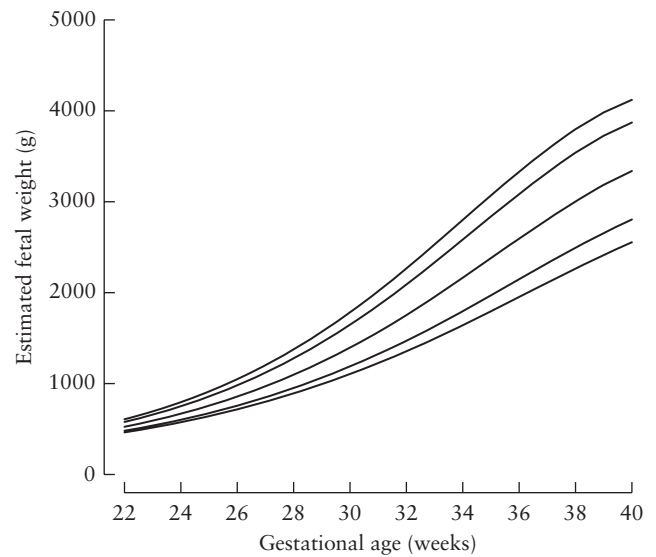

Figure 2 Smoothed 3rd, 10th, 50th, 90th and 97th centile curves for estimated fetal weight.

Table 2 Equations for parameters and computation of Z-scores and centiles for estimated fetal weight (EFW) in relation to gestational age (GA) in exact weeks

Parameter	Equation
Skewness	$\lambda(GA) = -4.257629 - 2162.234 \times GA^{-2} + 0.0002301829 \times GA^3$
Mean	$\mu(GA) = 4.956737 + 0.0005019687 \times GA^3 - 0.0001227065 \times GA^3 \times \log(GA)$
Coefficient of variation	$\sigma(GA) = 10^{-4} \times (-6.997171 + 0.057559 \times GA^3 - 0.01493946 \times GA^3 \times \log(GA))$
Z-score	$Y = \log(EFW)$ If $\lambda(GA) = 0$, $Z(GA) = \sigma(GA)^{-1} \times \log[Y/\mu(GA)]$ If $\lambda(GA) \neq 0$, $Z(GA) = [\sigma(GA) \times \lambda(GA)]^{-1} \times [(Y/\mu(GA))^{\lambda(GA)} - 1]$
Centiles	Z_α defined by $\Pr(z \leq Z_\alpha) = \alpha$ for $z \sim N(0,1)$, i.e. $Z_\alpha = \Phi^{-1}(\alpha)$ If $\lambda(GA) = 0$, $\log[C_\alpha(GA)] = \mu(GA) \times \exp[\sigma(GA) \times Z_\alpha]$ If $\lambda(GA) \neq 0$, $\log[C_\alpha(GA)] = \mu(GA) \times [Z_\alpha \times \sigma(GA) \times \lambda(GA) + 1]^{1/\lambda(GA)}$

the international standards, are presented in Figure 2. The corresponding equations for $\lambda(t)$, $\mu(t)$ and $\sigma(t)$, are presented in Table 2, allowing readers to calculate Z-scores. By estimating the EFW and knowing the gestational age, desired centiles can be calculated. For example, if AC = 26 cm and HC = 29 cm, at 30 + 0 weeks:

$$\log(EFW) = 5.084820 - 54.06633 \times (26/100)^3$$
$$- 95.80076 \times (26/100)^3 \times \log(26/100) + 3.136370$$
$$\times (29/100) = 7.312292.$$

Therefore, EFW = exp(7.312292) = 1499 g.

To compute the corresponding Z-score at 30 weeks' gestation, using the equations in Table 2 we must first calculate:

$$\lambda(30) = -4.257629 - 2162.234 \times 30^{-2}$$
$$+ 0.0002301829 \times 30^3 = -0.4451729.$$

$$\mu(30) = 4.956737 + 0.0005019687 \times 30^3$$
$$- 0.0001227065 \times 30^3 \times \log 30 = 7.241468.$$

$$\sigma(30) = 10^{-4} \times (-6.997171 + 0.057559 \times 30^3$$
$$- 0.01493946 \times 30^3 \times \log 30) = 0.017517.$$

Finally,

$$Z = (0.017517 \times -0.4451729)^{-1} \times$$
$$((7.312292/7.241468)^{-0.4451729} - 1) = 0.5617023.$$

Similarly, the 3rd centile ($\alpha = 0.03$), i.e. $Z = -1.88$ at 30 + 0 weeks, is calculated as follows using the equations in Table 2:

$$\log C_{0.03}(30) = 7.241468 \times (-1.88 \times 0.017517$$
$$\times (-0.4451729) + 1)^{-1/0.4451729} = 7.008552.$$

The 3rd centile for EFW at 30 weeks' gestation is therefore:

$$C_{0.03}(30) = \exp(7.008552) = 1106 \text{ g}.$$

The actual values for the 3rd, 10th, 50th, 90th and 97th centiles according to gestational age are presented in Table S1.
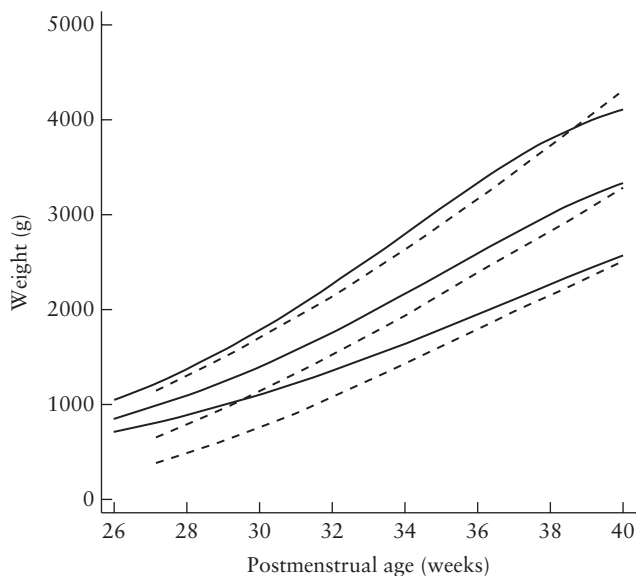
Figure 3 Comparison of fitted 3rd, 50th and 97th centiles for estimated fetal weight (——) with those of INTERGROWTH-21st preterm postnatal weight, with both sexes combined (- - -).

DISCUSSION

The INTERGROWTH-21st Project provides standards for early human growth based on populations that conform to the prescriptive approach recommended by the WHO[21,51]. By prescriptive, we mean that we observed a cohort of prospectively enrolled women whose risk of adverse maternal and perinatal outcomes (including fetal growth restriction) was low, based on their individual clinical profiles and the socioeconomic and demographic characteristics of the underlying geographically diverse populations. In fact, the INTERGROWTH-21st Project is unique because it has produced, for the first time, fetal ultrasound, newborn size and preterm postnatal growth datasets that have all been collected from the same underlying populations using the same rigorously applied methodologies.

We now present international EFW standards to complement the existing set, along with a formula for EFW based on HC and AC. Compared with several previous formulae[24], we found that FL did not improve the EFW, which agrees with previous work, in particular in growth-restricted fetuses[52]. Furthermore, it is likely that incorporating FL into the formula would increase the prediction error, as its measurement is associated with the highest inter- and intraobserver variability compared with AC and HC[53].

Unusually, we lowered the starting gestational age to 22 weeks, 2 weeks below the customary cut-off of 24 weeks' gestation for viability, for two reasons: to facilitate early recognition of fetal growth restriction around the recommended time of the second-trimester anatomy scan and to anticipate a possible extension of the limit of viability[54,55].

At the upper end of gestation, the centiles closely match those of the INTERGROWTH-21st Newborn Size Standards at 40 weeks[7]. The 3rd, 50th and 97th EFW

centiles at 40 weeks are 2554 g, 3338 g and 4121 g, respectively (Table S1), whereas for newborns (sexes combined) they are 2591 g, 3321 g and 4154 g, respectively (Figure S4). These similarities between fetal- and birth-weight centiles suggest that our model is valid for developing a formula for EFW using ultrasound biometry.

In contrast, there are significant discrepancies earlier in pregnancy (Figure 3). For example, at 33 weeks' gestation, the 3rd, 50th and 97th EFW centiles are 1495 g, 1954 g and 2529 g, respectively (Table S1); for newborns (sexes combined), they are 1190 g, 1903 g and 2715 g, respectively. It is possible that these differences are due to an overrepresentation of small, as well as, to a lesser extent, large babies in preterm births, even in the selected pregnant and newborn populations we studied.

The EFW formula and standards we present are also unique because we avoided the many common limitations identified by previous reviews[5,24]: retrospective design; use of routinely obtained measurements; suboptimal pregnancy dating strategies; variable time-to-birth without controlling for bias; absence of prospective ultrasound quality control, standardization and calibration of equipment; hospital-based sampling; absence of sampling from a healthy, well-nourished, underlying population; and no blinding of measurements.

Conversely, our standards are prescriptive, whereas reference charts describe only fetal size at a given place and time. The standards were derived prospectively, population-based and multinational. We have shown (using several analytical strategies) that the eight populations were consistently similar and could be pooled to create international standards[51]. Uniform research methods, protocols, processes and measurement tools were used throughout; these were combined with standardized identical equipment, training, a centralized electronic data management system and close monitoring of staff. The analytical approach aimed at identifying and correcting potential biases, and followed WHO recommendations to present the observed and smoothed data and explore the best fitting model with an *a-priori* strategy[56].

Using ultrasound, we examined separately HC, AC and FL, providing a comprehensive evaluation of structures that have different growth patterns; these measurements are often combined to calculate EFW. There are advantages in using a summary approximation: it is the most commonly measured marker of size at birth; as birth weight is associated with morbidity and mortality, it is helpful when counseling parents and enables pediatricians to make management decisions[57]; it may also help to refine the management of large babies.

However, there are also disadvantages in using only a single summary measure of size: first, there is a loss of the most granular information available when using the individual measurements, in terms of fetal skeletal and fat-based growth. Second, the fact that the individual measurement errors are compounded means that estimation is prone to inaccuracy; previous studies have shown that 95% prediction intervals for random error are in the region of ±14% of birth weight, and this is a particular

problem in low- and high-birth-weight babies[24]. Finally, as for other ultrasound measurements, there are numerous locally-derived EFW equations and reference charts[24] but, until now, no international standards existed, unlike the situation for newborn size and infant growth[7,8,10]. This may be, at least partly, responsible for the poor efficiency of screening strategies using biometry and EFW[58].

Therefore, we strongly recommend that, for clinical use, all individual fetal measurements, together with the summary measure of EFW, should be used together to make clinical decisions. In perinatal medicine, there is no room for a quick, minimalist approach that might lead to the early delivery of an at-risk fetus. Finally, implementation of the standards may raise concerns regarding the generalizability of data originating from a limited number of sites and/or a highly selected, low-risk population. As we have argued previously[11], having separate standards for a given country, institution or ethnic group has no biological basis and makes little sense in modern, multicultural societies. The international INTERGROWTH-21st standards describe optimal growth and can be used to assess both individuals and populations.

ACKNOWLEDGMENTS

REFERENCES

1. de Onis M, Habicht J-P. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *Am J Clin Nutr* 1996; **64**: 650–658.
2. Bakalis S, Peeva G, Gonzalez R, Poon LC, Nicolaides KH. Prediction of small-for-gestational-age neonates: screening by biophysical and biochemical markers at 30–34 weeks. *Ultrasound Obstet Gynecol* 2015; **46**: 446–451.
3. Lesmes C, Gallo DM, Gonzalez R, Poon LC, Nicolaides KH. Prediction of small-for-gestational-age neonates: screening by maternal serum biochemical markers at 19–24 weeks. *Ultrasound Obstet Gynecol* 2015; **46**: 341–349.
4. Conde-Agudelo A, Papageorghiou AT, Kennedy SH, Villar J. Novel biomarkers for predicting intrauterine growth restriction: a systematic review and meta-analysis. *BJOG* 2013; **120**: 681–694.
5. 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; **119**: 1425–1439.
6. 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; **384**: 869–879.
7. Villar J, Ismail LC, Victora CG, Ohuma EO, Bertino E, Altman DG, Lambert A, Papageorghiou AT, Carvalho M, Jaffer YA, Gravett MG, Purwar M, Frederick IO, Noble AJ, Pang R, Barros FC, Chumlea C, Bhutta ZA, Kennedy SH. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014; **384**: 857–868.
8. Villar J, Giuliani F, Bhutta ZA, Bertino E, Ohuma EO, Ismail LC, Barros FC, Altman DG, Victora C, Noble JA, Gravett MG, Purwar M, Pang R, Lambert A, Papageorghiou AT, Ochieng R, Jaffer YA, Kennedy SH; International Fetal and Newborn Growth Consortium for the 21(st) Century (INTERGROWTH-21(st)). Postnatal growth standards for preterm infants: the Preterm Postnatal Follow-up Study of the INTERGROWTH-21(st) Project. *Lancet Glob Health* 2015; **3**: e681–691.
9. 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.
10. Onis M de, Garza C, Onyango AW, Martorell R. WHO — Acta Paediatrica supplement on the WHO Child Growth Standards. WHO. http://www.who.int/childgrowth/standards/acta&uscore;supplement/en/. Published 2006.
11. Villar J, Papageorghiou AT, Pang R, Salomon LJ, Langer A, Victora C, Purwar M, Chumlea C, Qingqing W, Scherjon SA, Barros FC, Carvalho M, Altman DG, Giuliani F, Bertino E, Jaffer YA, Cheikh Ismail L, Ohuma EO, Lambert A, Noble JA, Gravett MG, Bhutta ZA, Kennedy SH. Monitoring human growth and development: a continuum from the womb to the classroom. *Am J Obstet Gynecol* 2015; **213**: 494–499.
12. Villar J, Papageorghiou AT, Pang R, Ohuma EO, Cheikh Ismail L, Barros FC, Lambert A, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Altman DG, Purwar M, Frederick IO, Noble JA, Victora CG, Bhutta ZA, Kennedy SH; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: the Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. *Lancet Diabetes Endocrinol* 2014; **2**: 781–792.

13. 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.

14. Wanyonyi SZ, Napolitano R, Ohuma EO, Salomon LJ, Papageorghiou AT. Image-scoring system for crown–rump length measurement. *Ultrasound Obstet Gynecol* 2014; **44**: 649–654.

15. Robinson HP, Fleming JE. A critical evaluation of sonar "crown–rump length" measurements. *Br J Obstet Gynaecol* 1975; **82**: 702–710.

16. 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. *Ultrasound Obstet Gynecol* 2013; **41**: 641–648.

17. Ohuma EO, Hoch L, Cosgrove C, Knight HE, Cheikh Ismail L, Juodvirsiene L, Papageorghiou AT, Al-Jabri H, Domingues M, Gilli P, Kunnawar N, Musee N, Roseman F, Carter A, Wu M, Altman DG; International Fetal and Newborn Growth Consortium for the 21st Century. Managing data for the international, multicentre INTERGROWTH-21st Project. *BJOG* 2013; **120** (Suppl 2): 64–70.

18. 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.

19. Salomon LJ, Bernard JP, Duyme M, Doris B, Mas N, Ville Y. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. *Ultrasound Obstet Gynecol* 2006; **27**: 34–40.

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**: 681–687.

21. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. *Food Nutr Bull* 2004; **25** (1 Suppl): S27–S36.

22. Cheikh Ismail L, Knight HE, Ohuma EO, Hoch L, Chumlea WC; International Fetal and Newborn Growth Consortium for the 21st Century. Anthropometric standardisation and quality control protocols for the construction of new, international, fetal and newborn growth standards: the INTERGROWTH-21st Project. *BJOG* 2013; **120** (Suppl 2): 48–55.

23. Cheikh Ismail L, Knight HE, Bhutta Z, Chumlea WC; International Fetal and Newborn Growth Consortium for the 21st Century. Anthropometric protocols for the construction of new international fetal and newborn growth standards: the INTERGROWTH-21st Project. *BJOG* 2013; **120** (Suppl 2): 42–47.

24. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 2005; **25**: 80–89.

25. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 1994; **43**: 429–467.

26. Kurmanavicius J, Burkhardt T, Wisser J, Huch R. Ultrasonographic fetal weight estimation: accuracy of formulas and accuracy of examiners by birth weight from 500 to 5000 g. *J Perinat Med* 2004; **32**: 155–161.

27. Cohen JM, Hutcheon JA, Kramer MS, Joseph KS, Abenhaim H, Platt RW. Influence of ultrasound-to-delivery interval and maternal–fetal characteristics on validity of estimated fetal weight. *Ultrasound Obstet Gynecol* 2010; **35**: 434–441.

28. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. CRC Press, 1994.

29. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc Ser C Appl Stat* 1999; **48**: 313–329.

30. Bland JM, Altman DG. Statistics Notes: Bootstrap resampling methods. *BMJ* 2015; **350**: h2622.

31. Ambler G, Royston P. Fractional polynomial model selection procedures: investigation of Type I error rate. *J Stat Comput Simul* 2001; **69**: 89–108.

32. Bellera CA, Hanley JA. A method is presented to plan the required sample size when estimating regression-based reference limits. *J Clin Epidemiol* 2007; **60**: 610–615.

33. Royston P. Constructing time-specific reference ranges. *Stat Med* 1991; **10**: 675–690.

34. Altman DG, Ohuma EO; International Fetal and Newborn Growth Consortium for the 21st Century. Statistical considerations for the development of prescriptive fetal and newborn growth standards in the INTERGROWTH-21st Project. *BJOG* 2013; **120** (Suppl 2): 71–76.

35. Altman DG, Chitty LS. Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; **3**: 378–384.

36. Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; **6**: 307–312.

37. Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals. *J R Stat Soc Ser A Stat Soc* 1997; **160**: 47–69.

38. Hynek M. Approaches for constructing age-related reference intervals and centile charts for fetal size. *Eur J Biomed Inform* 2010; **6**: 51–60.

39. Cole TJ. Fitting smoothed centile curves to reference data. *J R Stat Soc Ser A Stat Soc* 1988; **151**: 385–418.

40. Cole TJ. Using the LMS method to measure skewness in the NCHS and Dutch National height standards. *Ann Hum Biol* 1989; **16**: 407–419.

41. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; **11**: 1305–1319.

42. Rigby RA, Stasinopoulos DM. Using the Box–Cox t distribution in GAMLSS to model skewness and kurtosis. *Stat Model* 2006; **6**: 209–229.

43. Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Stat Med* 2004; **23**: 3053–3076.

44. Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press, 1993.

45. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci* 1996; **11**: 89–121.

46. van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med* 2001; **20**: 1259–1277.

47. Wade A, Kurmanavicius J. Creating unbiased cross-sectional covariate-related reference ranges from serial correlated measurements. *Biostatistics* 2009; **10**: 147–154.

48. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015. https://www.R-project.org.

49. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *J R Stat Soc Ser C Appl Stat* 2005; **54**: 507–554.

50. Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Software* 2007; **23**: 1–46.

51. Garza C, de Onis M. Rationale for developing a new international growth reference. *Food Nutr Bull* 2004; **25** (1 Suppl): S5–S14.

52. Proctor LK, Rushworth V, Shah PS, Keunen J, Windrim R, Ryan G, Kingdom J. Incorporation of femur length leads to underestimation of fetal weight in asymmetric preterm growth restriction. *Ultrasound Obstet Gynecol* 2010; **35**: 442–448.

53. 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**: 266–273.

54. Salomon LJ, Alfirevic Z, Berghella V, Bilardo C, Hernandez-Andrade E, Johnsen SL, Kalache K, Leung KY, Malinger G, Munoz H, Prefumo F, Toi A, Lee W; ISUOG Clinical Standards Committee. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol* 2011; **37**: 116–126.

55. Pignotti MS, Donzelli G. Perinatal care at the threshold of viability: an international comparison of practical guidelines for the treatment of extremely preterm births. *Pediatrics* 2008; **121**: e193–198.

56. Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, Van Buuren S, Pan H, Molinari L, Martorell R, Onyango AW, Martines JC. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Stat Med* 2006; **25**: 247–265.

57. Medchill MT, Peterson CM, Kreinick C, Garbaciak J. Prediction of estimated fetal weight in extremely low birth weight neonates (500–1000 g). *Obstet Gynecol* 1991; **78**: 286–290.

58. Stirnemann JJ, Benoist G, Salomon LJ, Bernard J-P, Ville Y. Optimal risk assessment of small-for-gestational-age fetuses using 31–34-week biometry in a low-risk population. *Ultrasound Obstet Gynecol* 2014; **43**: 311–316.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Appendix S1 Members of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st and INTERBIO-21st) and its Committees

Figure S1 Relationship between fetal weight and abdominal circumference in the final model, plotted for a fixed head circumference of 26 cm.

Figure S2 Relationship between fetal weight and head circumference in the final model, plotted for a fixed abdominal circumference of 23 cm.

Figure S3 Bias in estimation of fetal weight as a function of time to birth, showing mean percent prediction error and 95% CI according to time between last ultrasound scan and birth.

Figure S4 Gestational age-specific centiles for estimated fetal weight (blue) and birth weight (red). 3rd, 10th, 50th, 90th and 97th centiles are shown.

Table S1 Estimated fetal weight per completed week of gestation at 3rd, 10th, 50th, 90th and 97th centiles