International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project

Aris T Papageorghiou, Eric O Ohuma, Douglas G Altman, Tullia Todros, Leila Cheikh Ismail, Ann Lambert, Yasmin A Jaffer, Enrico Bertino, Michael G Gravett, Manorama Purwar, J Alison Noble, Ruyan Pang, Cesar G Victora, Fernando C Barros, Maria Carvalho, Laurent J Salomon, Zulfiqar A Bhutta*, Stephen H Kennedy*, José Villar*, for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st)†

Summary

Background In 2006, WHO produced international growth standards for infants and children up to age 5 years on the basis of recommendations from a WHO expert committee. Using the same methods and conceptual approach, the Fetal Growth Longitudinal Study (FGLS), part of the INTERGROWTH-21st Project, aimed to develop international growth and size standards for fetuses.

Methods The multicentre, population-based FGLS assessed fetal growth in geographically defined urban populations in eight countries, in which most of the health and nutritional needs of mothers were met and adequate antenatal care was provided. We used ultrasound to take fetal anthropometric measurements prospectively from 14 weeks and 0 days of gestation until birth in a cohort of women with adequate health and nutritional status who were at low risk of intrauterine growth restriction. All women had a reliable estimate of gestational age confirmed by ultrasound measurement of fetal crown–rump length in the first trimester. The five primary ultrasound measures of fetal growth—head circumference, biparietal diameter, occipitofrontal diameter, abdominal circumference, and femur length—were obtained every 5 weeks (within 1 week either side) from 14 weeks to 42 weeks of gestation. The best fitting curves for the five measures were selected using second-degree fractional polynomials and further modelled in a multilevel framework to account for the longitudinal design of the study.

Findings We screened 13 108 women commencing antenatal care at less than 14 weeks and 0 days of gestation, of whom 4607 (35%) were eligible. 4321 (94%) eligible women had pregnancies without major complications and delivered live singletons without congenital malformations (the analysis population). We documented very low maternal and perinatal mortality and morbidity, confirming that the participants were at low risk of adverse outcomes. For each of the five fetal growth measures, the mean differences between the observed and smoothed centiles for the 3rd, 50th, and 97th centiles, respectively, were small: 2·25 mm (SD 3·0), 0·02 mm (3·0), and –2·69 mm (3·2) for head circumference; 0·83 mm (0·9), –0·05 mm (0·8), and –0·84 mm (1·0) for biparietal diameter; 0·63 mm (1·2), 0·04 mm (1·1), and –1·05 mm (1·3) for occipitofrontal diameter; 2·99 mm (3·1), 0·25 mm (3·2), and –4·22 mm (3·7) for abdominal circumference; and 0·62 mm (0·8), 0·03 mm (0·8), and –0·65 mm (0·8) for femur length. We calculated the 3rd, 5th 10th, 50th, 90th, 95th and 97th centile curves according to gestational age for these ultrasound measures, representing the international standards for fetal growth.

Interpretation We recommend these international fetal growth standards for the clinical interpretation of routinely taken ultrasound measurements and for comparisons across populations.

Funding Bill & Melinda Gates Foundation.

Introduction

Screening for disturbances in fetal growth is one of the main purposes of antenatal care. None of the biomarkers assessed so far can accurately predict fetal growth restriction;[1] therefore, screening relies on routine measurement of uterine fundal height, complemented by ultrasound measurement of fetal size in women with pregnancy complications or with a relevant history or clinical evidence of fetal growth restriction.

Despite the widespread use of ultrasound worldwide, concerns have been expressed about the low detection rates of abnormal fetal growth in routine practice,[2,3] even when used mostly in high-risk subpopulations. However, these observations should be interpreted with caution in view of the large number of locally derived reference charts available[4] and the absence of suitable international standards similar to the standards used for monitoring infant growth.[5] Additionally, large variation is seen in the cutoff points (eg, 3rd, 5th, or 10th centile) used to establish whether fetal growth is abnormal, even within the same population or region.[4] The use of such a range of charts and cutoff points[4,6] in clinical decision making about fetal growth patterns inevitably leads to diagnostic confusion, difficulties comparing outcomes across

Child Health, Aga Khan University, Karachi, Pakistan (Prof Z A Bhutta PhD); and Center for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada (Prof Z A Bhutta)

Correspondence to:
Prof José Villar, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Women's Centre, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
jose.villar@obs-gyn.ox.ac.uk

populations, and unnecessary anxiety for mothers and their families.

International standards are urgently needed because, unlike for the neonatal period, none are available to enable fetal growth to be monitored as recommended by WHO.[7] The aim of the Fetal Growth Longitudinal Study (FGLS)—part of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) Project—was to construct such standards for fetal growth, adopting the same approach and methods as the WHO Multicentre Growth Reference Study (MGRS)[8] by studying a cohort of healthy, well nourished, pregnant women from eight geographically diverse populations who were at low risk of adverse maternal and perinatal outcomes.[9,10]

Methods
Study design and participants
INTERGROWTH-21st is a multicentre, multiethnic, population-based project done between April 27, 2009, and March 2, 2014, in eight sites in eight countries: Pelotas, Brazil; Turin, Italy; Muscat, Oman; Oxford, UK; Seattle, WA, USA; Shunyi County, a suburban district of the Beijing municipality, China; the central area of the city of Nagpur (Central Nagpur), Maharashtra, India; and the Parklands suburb of Nairobi, Kenya.[9] The INTERGROWTH-21st Project's main aim was to study growth, health, nutrition and neurodevelopment from less than 14 weeks and 0 days of gestation to 2 years of age, using the same conceptual framework as the WHO MGRS,[8] so as to produce prescriptive growth standards to complement the existing WHO Child Growth Standards,[5] and to develop a new phenotypic classification of the fetal growth restriction and preterm birth syndromes.

FGLS is one of the three main studies of the INTERGROWTH-21st Project, which has been described in detail elsewhere.[9] Briefly, the populations that contributed participants to the project were first selected at the geographical level and then at the individual level within each study site. At the population level, we identified an urban area (a complete city or county, or part of a city with clear political or geographical limits) where most deliveries occurred in health facilities. The areas had to be located at an altitude of 1600 m or lower and women receiving antenatal care had to plan to deliver in these facilities or in a similar hospital located in the same geographical area. Furthermore, the area had to be free or have low levels of major, known, non-microbiological contamination such as pollution, domestic smoke due to tobacco smoking or cooking, radiation, or any other toxic substances.[11] At each of the eight study sites, we selected all facilities providing antenatal and intrapartum care where more than 80% of deliveries in those urban areas occurred.

At each study site we recruited women with the characteristics required for FGLS—ie, women with no clinically relevant obstetric, gynaecological, or medical history who initiated antenatal care before 14 weeks of gestation (measured by menstrual dates) and met the entry criteria of optimal health, nutrition, education, and socioeconomic status. A detailed description of the entry criteria and definitions has been published previously.[9] For example, adequate nutritional status was defined in the first trimester according to maternal height (≥153 cm), body-mass index (BMI; ≥18·5 and <30 kg/m²), and haemoglobin concentration (≥110 g/L), and whether the mother was receiving treatment for anaemia or following any special diets—eg, vegetarian with no animal products. This resulted in a group of educated, affluent, clinically healthy women, with adequate nutritional status, who by definition were at low risk of fetal growth restriction and preterm birth.

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 where the project was implemented. Participants provided written consent to be involved in the study.

Procedures
The last menstrual period (LMP) was used to calculate gestational age provided that the LMP date was certain; the woman had a regular 24–32 day menstrual cycle; 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 crown–rump length (CRL), measured by ultrasound between 9 weeks and 0 days and 13 weeks and 6 days after the LMP, was 7 days or less, using the formula described by Robinson and Fleming.[12] To ensure that CRL measures were interpreted consistently, the Robinson and Fleming formula was loaded into all the study ultrasound machines; whenever another machine had to be used locally for CRL measurement, a conversion table extracted from the same formula was provided. The CRL technique was also standardised across sites and all ultrasonographers were trained uniformly.[13]

We scanned women every 5 weeks (within 1 week either side) after the initial dating scan, so that the possible ranges after the dating scan were 14–18, 19–23, 24–28, 29–33, 34–38, and 39–42 weeks of gestation. At each visit, fetal head circumference, biparietal diameter, occipitofrontal diameter, abdominal circumference, and femur length were measured three times from three separately obtained ultrasound images of each structure. The detailed measurement protocol and the unique standardisation procedures have been reported elsewhere,[14] and all documentation, protocols, 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. 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. Callipers were then placed on the outer border of the parietal bones (outer to outer) at the widest or longest part of the skull for the biparietal diameter and occipitofrontal diameter, respectively; the head circumference was measured using the ellipse facility on the outer border of the skull.

The measurements of the abdomen were taken in a cross-sectional view of the fetal abdomen as close as possible to circular, with the umbilical vein in the anterior third of the abdomen (at the level of the portal sinus), with the stomach bubble visible. The operator 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, preferably, and the spine had to be at either a 3 o'clock 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.

Recognising that the head and abdomen are elliptical in shape, in addition to measurements of head circumference and abdominal circumference obtained using the ellipse facility of the ultrasound machine, head circumference and abdominal circumference were also estimated from the lengths of the two directly measured diameters (major and minor axes). Because no explicit (closed form) formula exists to calculate the circumference of an ellipse from the diameters, common approximations are used, based on the formula for a circle. These approximations tend to be more accurate when the two diameters are similar, especially when the ratio of biparietal diameter to occipitofrontal diameter is about 1; however, when this ratio is less than 0·8 it can lead to bias. Therefore, we calculated head circumference and abdominal circumference using an exact formula derived from the elliptical integral (appendix).

Finally, the femur length 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 visualised, unobscured by shadowing from adjacent bony parts, and the femur had to fill at least 30% of the monitor screen. The intersection of the callipers was placed on the outer borders of the edges of the femoral diaphysis (outer to outer) ensuring clear femoral edges; ultrasound artefacts 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 localised are available on the INTERGROWTH-21st website).

To reduce expected value bias, all measurements taken at or after 14 weeks and 0 days of gestation were not visible on the screen to the ultrasonographer, which was

achieved for the INTERGROWTH-21st Project by specially adapting the ultrasound machine used at all sites (Philips HD-9 [Philips Ultrasound, Bothell, WA, USA] with curvilinear abdominal transducers C5-2, C6-3, V7-3).[14] All ultrasound data were submitted electronically to the study's database. After each set of measurements, ultrasonographers scored the quality of their images on the basis of standard image-scoring criteria.[15,16] Images that did not score the maximum amount of points were repeated until the best possible score was achieved. The training, standardisation, and quality-control methods used across all sites are described in detail elsewhere.[15] In brief, ultrasonographers were recruited on the basis of their technical experience, motivation, reliability, and ability to speak the local language or languages. They underwent rigorous training, consisting of acquisition of theoretical knowledge and familiarity with the study protocol, ultrasound operations manual, ultrasound machine, data collection, and quality-control measures. Centralised hands-on training and initial standardisation were also done.[17] Additionally, site-specific standardisation was done at regular intervals by the ultrasound quality-control unit to ensure proper use of the ultrasound equipment, calibration, and adherence to the protocol. A quality-control system was implemented throughout the study with two main objectives: assessment of the distributions of the three blinded measurements for each fetal biometric measure at every scan and construction of cumulative sum charts; and assessment of the quality of a random 10% sample of all ultrasound images by the centralised ultrasound quality-control unit using a validated scoring system, and remeasurement of these images.[15] Only after three measurements of each structure were recorded were the average values revealed to the operator for clinical purposes.

All documentation used in the INTERGROWTH-21st Project was tested locally and introduced into the specially developed, online electronic data entry, cleaning, and management system developed by MedSciNet UK (London, UK). Data were entered locally directly onto the web-based system.[18] During data cleaning, we excluded 14 head circumference, six biparietal diameter, two occipitofrontal diameter, 14 abdominal circumference, and seven femur length measurements because they were regarded as implausible on the basis of all sites' gestational age distribution or were more than 5 SDs of all sites' gestational age-specific mean.

See **Online** for appendix

Statistical analysis
The sample size was based on pragmatic and statistical considerations; the latter focused on the precision and accuracy of one extreme centile—ie, the 3rd or 97th centile, and regression-based reference limits.[19,20] We have shown that a sample of 4000 women would obtain precision of 0·03 SD at the 3rd or 97th centile. Further details on the precision obtained at the 5th or 10th centiles by sample size (ranging from 500 to 6000) were

included in a table in a previous publication.[21] Our target was a sample of, on average, 500 pregnant women per study site, after excluding complicated pregnancies and those lost to follow-up. This sample size is larger than most previous longitudinal studies and was considered feasible and adequate to explore site-specific differences or subgroups of about 10% of the whole study if needed.

We expected that, overall, less than 3% of participants would be lost to follow-up, and that another 3% would be excluded (using criteria decided a priori) from the study population because of fetal deaths or congenital abnormalities. We also planned to exclude mothers diagnosed with very severe medical disorders (eg, cancer or HIV) during the index pregnancy, those with severe unanticipated pregnancy-related disorders needing hospital admission (eg, eclampsia or severe pre-eclampsia), and those identified during pregnancy who no longer fulfilled all the entry criteria (eg, women who started smoking during pregnancy or had an episode of malaria). The decision to pool the data from all the study sites to construct the fetal growth standards was based on the detailed analyses done using the strategy that was recommended in the WHO MGRS; the results are described in detail elsewhere.[10] The statistical methods used to construct the fetal standards from this pooled sample were based on the recommendations of Altman and Chitty[22] and Royston and Altman,[23] complemented by recent scientific literature reviews[24,25] and our systematic review of the methods used in previous ultrasound studies aimed at creating references of fetal size.[26] Our overall aim was to produce centiles that change smoothly with age and maximise simplicity without compromising model fit.

We explored the following statistical methods: mean and SD method using fractional polynomials;[27] Cole's lambda (λ), mu (μ), and sigma (σ) (LMS) method,[28–30] 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[31] (ie, lambda, mu, sigma, assuming Box-Cox t distribution) method, which assumes a shifted and scaled (truncated) t distribution to take account of skewness and leptokurtosis; the LMSP[32] (ie, lambda, mu, sigma, assuming Box-Cox power exponential distribution) method, which assumes a Box-Cox power exponential distribution to take account of skewness, platykurtosis, and leptokurtosis; and multi-level models[33,34] to account for repeat measurements. Furthermore, to present the curves, we assessed three smoothing techniques: fractional polynomials,[27] cubic splines,[35] and penalised splines.[36]

After comparing results from the various approaches, we found no evidence to support a non-normal distribution for a specific gestational age. Thus, more complex methods allowing for skewness and kurtosis were not needed and the following methods were chosen. The best fitting powers for the median head circumference, biparietal diameter, occipitofrontal diameter, abdominal circumference, and femur length were provided by second-degree fractional polynomials and further modelled in a multilevel framework to account for the longitudinal design of the study. The data structure is composed of three levels—ie, measurements within occasions within participants. We thus fitted a three-level hierarchical model using the runmlwin package in Stata.[34] To obtain an equation for the SD, we modelled the resulting variance components from the multilevel model that accounts for the between-participant and within-participant correlations using fractional polynomials. The SD was modelled on the log scale to stabilise variance.

Goodness of fit incorporated visual inspection of overall model fit by comparing empirical centiles (calculated per completed week of gestation—eg, 38 weeks of gestation is equal to 38 weeks and 0 days to 38 weeks and 6 days of gestation) to the fitted centiles, using quantile-quantile (q-q) plot of the residuals, plots of residuals versus fitted values, and the distribution of fitted Z scores across gestational ages. Adjusting for correlated data within fetuses has a well known effect of making fitted centiles slightly narrower by reducing random error;[37,38] therefore, we would expect the fitted centiles to be slightly narrower than empirical ones. All analyses were done in R statistical software[39] using the Generalised Additive Models for Location, Scale and Shape (GAMLSS) framework[40,41] and Stata, version 11.2, software.

Tables containing mean and SD, centile values, and Z scores by gestational age, expressed in completed weeks of gestation (as recommended by WHO International Statistical Classification of Diseases and Related Health Problems 10th Revision), as well as printable charts will be available free on the INTERGROWTH-21st website. A tool to calculate the individual centiles and Z scores by gestational age (in exact weeks and days) will also be available free on the same website by December, 2014.

Role of the funding source

Results
We screened 13 108 pregnant women attending the study clinics at less than 14 weeks and 0 days of gestation within the project's defined geographical areas, of whom 4607 (35%) met the eligibility criteria, provided consent, and were enrolled (figure 1). The most common reasons for ineligibility were low maternal height (1022 [12%] of 8501 ineligible patients), BMI 30 or higher (1009 [12%]), and maternal age younger than 18 years or older than 35 years (915 [11%]).

The contribution of each site to the total study population ranged from 311 (7%) of 4607 in the USA to 640 (14%) in the UK. A detailed description of each site's population and their characteristics have been presented before.[9,10] 36 (<1%) women who developed severe disorders during pregnancy or took up smoking or used drugs were excluded, and 71 (2%) were lost to follow-up or withdrew consent. 4422 (96%) women delivered a live singleton, of whom 4321 (98%) were born without congenital malformations; their longitudinal fetal growth data were used to construct the fetal standards reported here.

A detailed description of the environmental characteristics and working conditions in each geographical area have been presented elsewhere.[11] The results of a survey using a data collection form specifically developed for the project showed that most women were not exposed to environmental hazards (eg, pollution, domestic smoke due to tobacco smoking or cooking, radiation, or any other toxic substances) that could affect pregnancy outcomes; housing conditions were as expected for relatively affluent populations in these communities in terms of size, water, sanitation, and construction characteristics.[11]
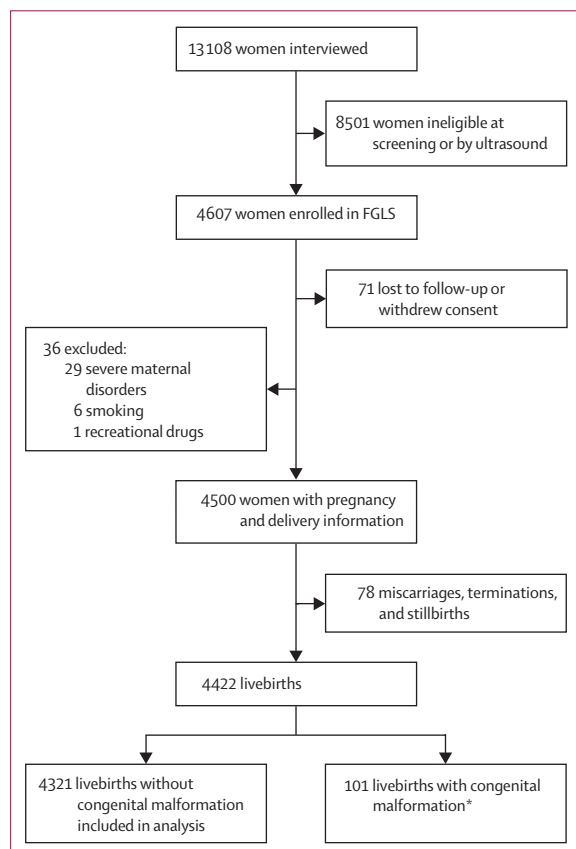
The baseline characteristics of the study cohort across the eight sites were very similar, which was expected because women were selected from the underlying low-risk populations using the same clinical and demographic criteria.[10] Table 1 shows the pregnancy and perinatal events for the complete cohort, which confirmed their status as healthy, well nourished women at low risk of impaired fetal growth.

The median number of ultrasound scans (excluding the dating scan) in all women was 5·0 (range 1–7; mean 4·9 [SD 0·8]) and 3976 (92%) women had four or more scans; of these 3976 women, the mean number of ultrasound scans was 5·0 (SD 0·6; range 4–7), suggesting that participants adhered well to the protocol. 17 261 (85%) of the 20 313 ultrasound scans were done within the expected gestational age window of the protocol (ranging from 2060 [76%] of 2707 in India to 2803 [93%] of 3006 in Oman).

The gestational age-specific observed and smoothed centiles for head circumference, biparietal diameter, occipitofrontal diameter, abdominal circumference, and

Figure 1: Fetal Growth Longitudinal Study (FGLS) flow chart
*Congenital malformations diagnosed by ultrasound during pregnancy or at birth by clinical examination.

	Pregnancies (n=4321)
Maternal age, years	28·4 (3·9)
Maternal height, cm	162·2 (5·8)
Maternal weight, kg	61·3 (9·1)
Paternal height, cm	174·4 (7·3)
Body-mass index, kg/m²	23·3 (3·0)
Gestational age at first visit, weeks	11·8 (1·4)
Years of formal education, years	15·0 (2·8)
Haemoglobin level at <15 weeks, g/L	125 (11)
Married or cohabiting	4204 (97%)
Nulliparous	2955 (68%)
Pre-eclampsia	31 (<1%)
Pyelonephritis	16 (<1%)
Any sexually transmitted infection	3 (<1%)
Spontaneous initiation of labour	2868 (66%)
PPROM (<37 weeks)	80 (2%)
Caesarean section	1541 (36%)
NICU admission >1 day	240 (6%)
Preterm (<37 weeks)	195 (5%)
Preterm and spontaneous onset of labour	126 (3%)
Term LBW (<2500 g; ≥37 weeks)	128 (3%)
Neonatal mortality	7 (<1%)
Male sex	2149 (50%)
Exclusive breastfeeding at discharge	3786 (88%)
Mother admitted to intensive care unit	17 (<1%)
Birthweight (≥37 weeks), kg	3·3 (0·4)
Birth length (≥37 weeks), cm	49·4 (1·9)
Birth head circumference (≥37 weeks), cm	33·9 (1·3)

Data are mean (SD) or number (%). Maternal baseline characteristics were measured at less than 14 weeks of gestation. PPROM=preterm prelabour rupture of membranes. NICU=neonatal intensive care unit. LBW=low birthweight.

Table 1: Baseline characteristics and perinatal events of the Fetal Growth Longitudinal Study
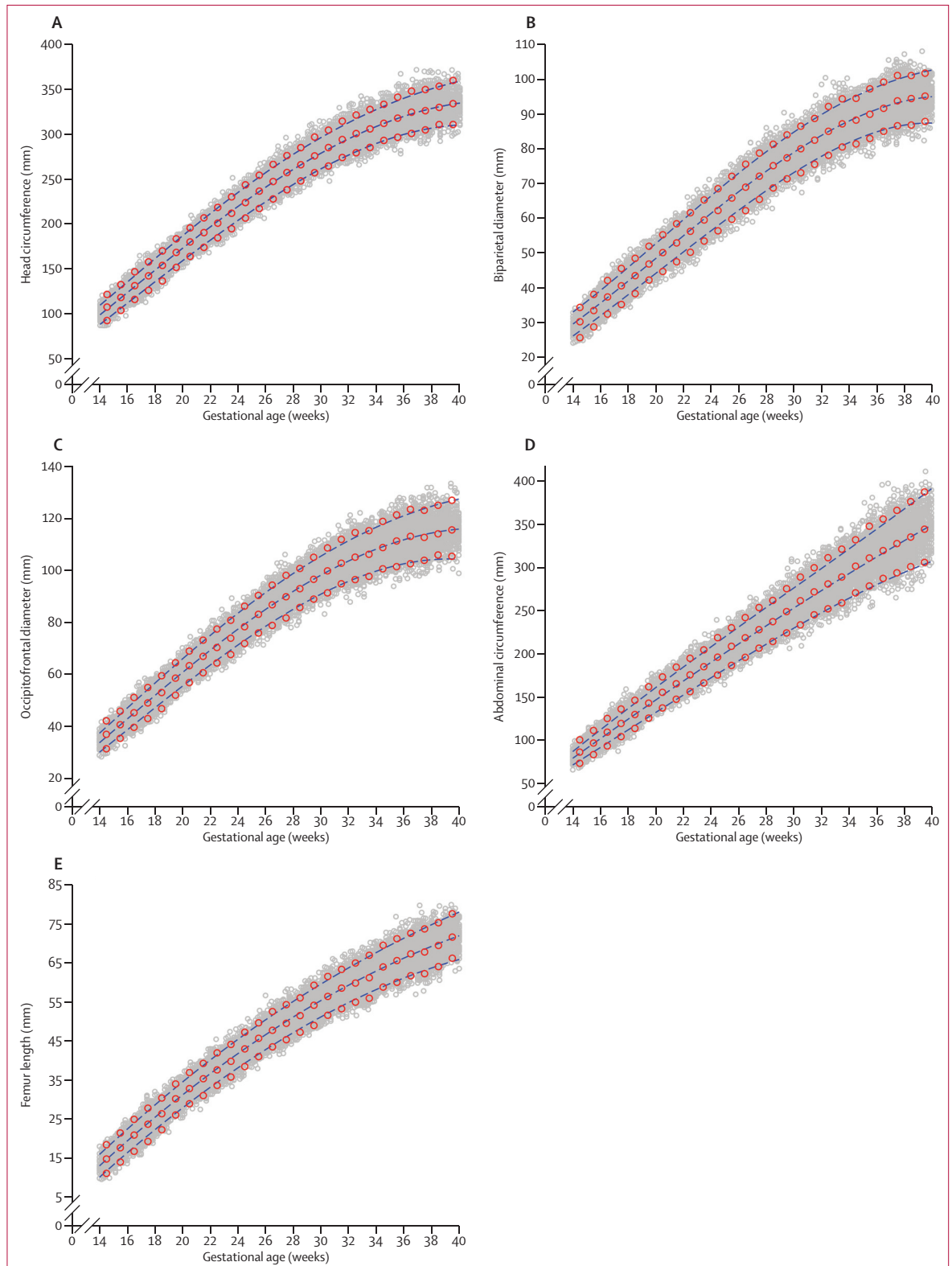
***Figure 2*: Fitted 3rd, 50th, and 97th smoothed centile curves of fetal measurements**
Fitted 3rd (bottom dashed line), 50th (middle dashed line), and 97th (top dashed line) smoothed centile curves for fetal head circumference (A), fetal biparietal diameter (B), fetal occipitofrontal diameter (C), fetal abdominal circumference (D), and fetal femur length (E) measured by ultrasound according to gestational age. Open red circles show empirical values for each week of gestation and open grey circles show actual observations.
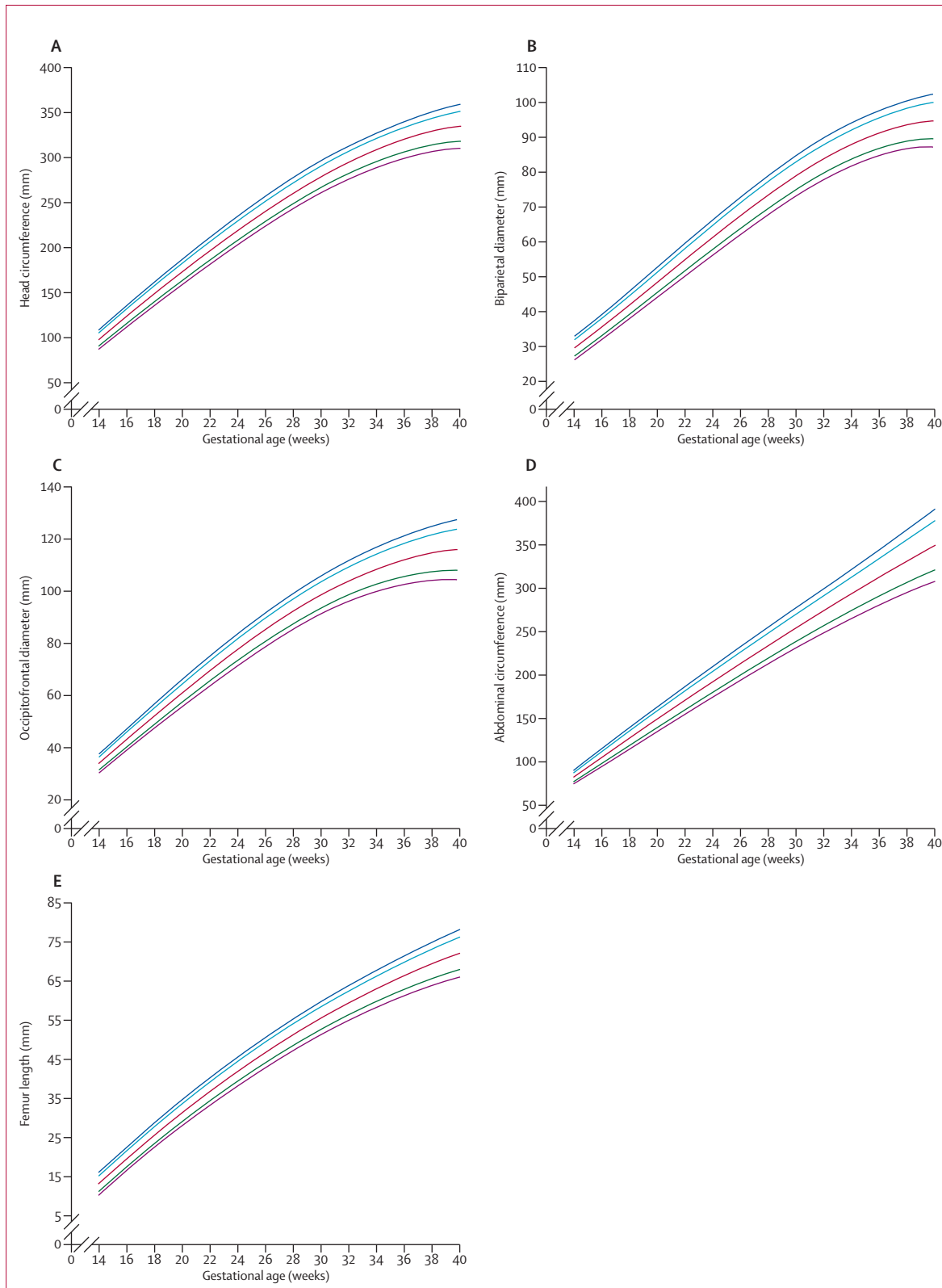
Figure 3: **3rd, 10th, 50th, 90th and 97th smoothed centile curves**
Fetal head circumference (A), fetal biparietal diameter (B), fetal occipitofrontal diameter (C), fetal abdominal circumference (D), and fetal femur length (E) measured by ultrasound according to gestational age.

femur length are presented in figure 2. Similarities between smoothed centile curves (3rd, 50th, and 97th centiles) and observed values were assessed by gestational age-specific comparisons. The comparisons show excellent agreement between the smoothed and empirical centiles.

Overall, the mean differences between smoothed and observed centiles for the 3rd, 50th, and 97th centiles, respectively, were small: 2·25 mm (SD 3·0), 0·02 mm (3·0), and −2·69 mm (3·2) for head circumference (figure 2A); 0·83 mm (0·9), −0·05 mm (0·8), and −0·84 mm (1·0) for biparietal diameter (figure 2B); 0·63 mm (1·2), 0·04 mm (1·1), and −1·05 mm (1·3) for occipitofrontal diameter (figure 2C); 2·99 mm (3·1), 0·25 mm (3·2), and −4·22 mm (3·7) for abdominal circumference (figure 2D); and 0·62 mm (0·8), 0·03 mm (0·8), and −0·65 mm (0·8) for femur length (figure 2E).

The 3rd, 5th, 10th, 50th, 90th, 95th, and 97th fitted centile curves for head circumference, biparietal diameter, occipitofrontal diameter, femur length, and abdominal circumference according to gestational age, which represent the international standards for fetal growth, are presented in figure 3. The corresponding equations for the mean and SD from the multilevel regression models for each measure are presented in table 2, allowing for calculations by readers of any desired centiles according to gestational age in exact weeks. For example, centiles can be calculated as mean $\pm Z \times$ SD, where Z is −1·88, −1·645, −1·28, 0, 1·28, 1·645, and 1·88 for the 3rd, 5th, 10th, 50th, 90th, 95th, and 97th centiles, respectively. The actual values for these centiles according to gestational age are presented in the appendix.

Regression equation
Head circumference
Mean	$-28\cdot2849 + 1\cdot69267 \times GA^2 - 0\cdot397485 \times GA^2 \times \log(GA)$
SD	$1\cdot98735 + 0\cdot0136772 \times GA^3 - 0\cdot00726264 \times GA^3 \times \log(GA) + 0\cdot000976253 \times GA^3 \times \log(GA)^2$
Biparietal diameter
Mean	$5\cdot60878 + 0\cdot158369 \times GA^2 - 0\cdot00256379 \times GA^3$
SD	$\exp(0\cdot101242 + 0\cdot00150557 \times GA^3 - 0\cdot000771535 \times GA^3 \times \log(GA) + 0\cdot0000999638 \times GA^3 \times \log(GA)^2)$
Occipitofrontal diameter
Mean	$-12\cdot4097 + 0\cdot626342 \times GA^2 - 0\cdot148075 \times GA^2 \times \log(GA)$
SD	$\exp(-0\cdot880034 + 0\cdot0631165 \times GA^2 - 0\cdot0317136 \times GA^2 \times \log(GA) + 0\cdot00408302 \times GA^2 \times \log(GA)^2)$
Abdominal circumference
Mean	$-81\cdot3243 + 11\cdot6772 \times GA - 0\cdot000561865 \times GA^3$
SD	$-4\cdot36302 + 0\cdot121445 \times GA^2 - 0\cdot0130256 \times GA^3 + 0\cdot00282143 \times GA^3 \times \log(GA)$
Femur length
Mean	$-39\cdot9616 + 4\cdot32298 \times GA - 0\cdot0380156 \times GA^2$
SD	$\exp(0\cdot605843 - 42\cdot0014 \times GA^{-2} + 0\cdot00000917972 \times GA^3)$

All log are natural logarithms. GA=exact gestational age.

Table 2: Equations for the estimation of the mean and SD (in mm) of each fetal biometry measurement according to exact gestational age (in weeks)

Discussion

To describe optimal fetal growth for singleton pregnancies, the INTERGROWTH-21st Project used a prescriptive approach[42] to select a population of healthy, well nourished pregnant women and their fetuses and newborn babies, following the recommendation of the special 1995 WHO Expert Committee[43] and implemented by the WHO MGRS for infants and children.[7] This meant studying a cohort of prospectively enrolled women whose risk of adverse maternal and perinatal outcomes (including fetal growth restriction) was low according to their individual clinical profiles and the socioeconomic and demographic characteristics of the underlying eight geographically diverse populations.

From this population, we have generated the first international standards (as opposed to references) for fetal growth using the complete FGLS sample of 4321 live singletons without congenital malformations, whose healthy status has been confirmed by close follow-up of the pregnancies, assessment of the neonates, and in accordance with examination of their growth, nutrition, and gross motor development at age 1 year.

Remarkably, at birth and age 1 year, the infants included in the international fetal growth standards presented here were very similar to those enrolled in the WHO MGRS, whose data were used to construct the infant and child growth standards that have now been adopted by more than 140 countries. For example, the mean birth length at more than 37 weeks and 0 days of gestation in the WHO MGRS was 49·5 cm (SD 1·9), compared to 49·4 cm (1·9) in the INTERGROWTH-21st cohort, and the mean birthweight in the WHO MGRS was 3·3 kg (0·5), which was virtually identical to ours 3·3 kg (0·4). Moreover, very reassuringly, at age 1 year, the FGLS cohort was on the 49th and 52nd centiles for length and the 49th and 50th centiles for head circumference (for boys and girls, respectively) of the WHO Child Growth Standards.

Therefore, the international fetal growth standards presented here contribute some crucial components to the care of pregnant women and their children: a unique set of clinical tools for use across all health-care systems to diagnose fetal growth restriction uniformly and monitor growth from early pregnancy, through the neonatal period to age 5 years, founded on the same conceptual and methodological approach and age-specific cutoff points.

We believe these standards, as opposed to the several locally produced references in use worldwide,[4] have the potential to improve pregnancy outcomes,[44] not least because at present the diagnosis of fetal growth restriction is made at different levels of care, even within the same regions or countries, using different fetal growth charts and cutoff points—ie, fetuses can be classified as growth restricted in one part of a city or country and of healthy size in another. This leads to inaccuracy in diagnosis and ultimately unnecessary, or an absence of, appropriate interventions. Additionally,

use of fetal growth standards derived from a healthy population reduces the risk of underdiagnosing fetal growth restriction, which can occur when the fetus is monitored against references that include high-risk mothers (panel).

The existing fetal growth references have some major limitations, which were identified in a systematic review.[4] They include retrospective design; use of routinely obtained measurements; suboptimal pregnancy dating strategies; absence of prospective ultrasound quality control, standardisation, and calibration of equipment; hospital-based sampling; absence of sampling from a healthy, well nourished, underlying population; and no blinding of measurements.[4] In view of these limitations and the reference nature of their design (ie, they are related to a given place and time), the identification of fetuses with growth restriction is, not surprisingly, problematic. Additionally, the choice of reference chart to be used is often based on local preference, or simply on the chart that is loaded into the ultrasound machine software. However, the choice has major clinical implications. For example, the proportion of fetuses classified as having a biparietal diameter below the 5th centile at 20–24 weeks of gestation ranged from 6·6% to 23·7% when using three different popular ultrasound references.[46]

Our study[9] aimed to avoid these limitations.[4] It was truly prospective, and we used a population-based sampling strategy that initially selected geographical regions where a substantial proportion of women are healthy, well nourished, educated, and at low risk of fetal growth restriction, from which, in a second step, we identified pregnant women for FGLS, on the basis of several individual clinical and demographic characteristics. The ultrasound measurements were taken specifically for the purpose of constructing international standards and the method was rigorous and implemented across all study sites: we standardised the measurement of fetal size using centrally trained staff; each study site used the same specially adapted ultrasound equipment to allow blinding of measurements; more than one measurement for each biometric variable was taken at each scan to reduce error; and we developed a novel quality-control strategy for all ultrasound measurements, including assessment of intraobserver and interobserver variability at all sites and continual independent image review and scoring at a central location where all images were stored for independent re-measuring.[14,15] Finally, the appropriate statistical methods were used to analyse our dataset of repeated measurements; hence, we are in effect describing true growth.

Unfortunately, not all eligible women from the underlying low-risk population were enrolled because of the logistical constraints of having only one dedicated ultrasound machine at each site for the study. Eligible women were recruited consecutively at each antenatal clinic in the geographical areas up to a weekly limit (about six women per week per site) to avoid

---

Panel: Research in context

Systematic review

Ultrasound is widely used worldwide to detect abnormal fetal growth, but its accuracy as a test is dependent upon the reliability of the fetal size chart used. A systematic review identified 83 observational studies, published from Jan 1, 1968, to Sept 30, 2011, whose primary aim was to create such charts. We searched Medline, Embase, and CINAHL for studies published in English. A list of search terms used has been published previously.[4] We found 337 studies, which had substantial methodological heterogeneity, and none met the WHO recommendations for monitoring fetal anthropometric measurements.[7] The large number of charts available, the reference nature of their design (ie, they are related to a given place and time), and the variation in the cutoff points used to define abnormal growth, make it difficult in routine clinical practice to identify fetal growth restriction (FGR). International standards are, therefore, urgently needed, in keeping with those available for children from birth to age 5 years.[5,44]

Interpretation

From our cohort of healthy, well nourished pregnant women prospectively enrolled from eight geographically diverse populations, and whose risk of adverse maternal and perinatal outcomes (including FGR) was low, we have generated the first international standards for fetal growth based on the primary ultrasound measures of head circumference, biparietal diameter, occipitofrontal diameter, abdominal circumference, and femur length, according to gestational age. Combined with our newborn standards,[45] the new fetal standards are for use worldwide to diagnose FGR uniformly and monitor growth from early pregnancy through to the neonatal period. We recommend these tools for the interpretation of routine ultrasound measurements and comparisons across populations.

---

overwhelming the capacity of the project's ultrasound research team. We believe that the risk of selection bias was small, and comparisons show that the FGLS cohort was very similar to the corresponding total population of women with the same characteristics from which the FGLS cohort was obtained.[10]

The inevitable and recurrent question related to the implementation of international, prescriptive growth standards (including ours) is whether or not they can be generalised to all populations, considering that they were generated, in our case, from only eight geographically diverse study sites. The answer is that the generalisability of anthropometric standards based on a prescriptive approach and international sampling frames of geographically and ethnically diverse populations is supported by the uncertainty surrounding the identification of functionally significant, common genetic variants that are unique to ethnic groups in quantitative, complex traits.[47,48] Additionally, we used a population-based sample of well nourished participants at each of the INTERGROWTH-21st sites—ie, we intentionally controlled for well documented adverse environmental or socioeconomic conditions that are universal constraints on fetal growth, thereby considerably reducing a common cause for population-wide growth differences often observed in diverse populations. Furthermore, the empirical data produced from the same population showed the similarity from early pregnancy to birth of all the fetal and newborn skeletal measures across the study sites.[10]

Finally, the cumulative experience gained from implementation in more than 140 countries of the WHO MGRS standards, which were derived from the prescriptive, international study of postnatal growth, and whose design and methods were almost identical to the INTERGROWTH-21st Project, further supports the generalisability of our standards. International growth standards can be regarded as close to universal as possible and more accurate than other available methods.

We hope that these international fetal growth standards, complementing our published standards for early fetal linear size,[49] in addition to our newborn standards,[45] will contribute to the recently launched Every Newborn Action Plan that focuses on the "development of a minimum perinatal dataset with standard metrics for counting births, stillbirths, neonatal deaths, birthweight, and gestational age".[50]

Contributors
JV and SHK conceptualised and designed the INTERGROWTH-21st Project. JV, SHK, DGA, and JAN prepared the original protocol, with input from ATP, LCI, FCB, and ZAB. JV, ATP, LI, AL, and ZAB supervised and coordinated the project's overall undertaking. EOO, DGA, ATP, and JV were responsible for data management and analysis in collaboration with CGV. RP, FCB, MC, YAJ, EB, MGG, MP, and TT collaborated in the overall project and implemented it in their respective countries. LJS was the senior adviser of the quality-control unit of the project. ATP, SK, and JV wrote the report with input from all authors. All authors read the report and made suggestions on its content.

Declaration of interests
ATP has received personal fees from Intelligent Ultrasound and non-financial support from Philips Ultrasound. MGG has received a sub-contract for the INTERGROWTH-21st Project from Oxford University. JAN has received personal fees from Intelligent Ultrasound and grants and non-financial support from Philips Ultrasound. All other authors declare no competing interests.

References
1 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:** 681–94.

2 Bricker L, Neilson JP, Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst Rev* 2008; **4:** CD001451.

3 Sylvan K, Ryding EL, Rydhstroem H. Routine ultrasound screening in the third trimester: a population-based study. *Acta Obstet Gynecol Scand* 2005; **84:** 1154–58.

4 Ioannou C, Talbot K, Ohuma E, et al. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. *BJOG* 2012; **119:** 1425–39.

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

6 Napolitano R, Dhami J, Ohuma E, et al. Pregnancy dating by fetal crown-rump length: a systematic review of charts. *BJOG* 2014; **121:** 556–65.

7 de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *Am J Clin Nutr* 1996; **64:** 650–58.

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

9 Villar J, Altman DG, Purwar M, et al. The objectives, design and implementation of the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 9–26.

10 Villar J, Papageorghiou AT, Pang R, et al. 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; published online July 7. http://dx.doi.org/10.1016/S2213-8587(14)70121-4.

11 Eskenazi B, Bradman A, Finkton D, et al. A rapid questionnaire assessment of environmental exposures to pregnant women in the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 129–38.

12 Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. *Br J Obstet Gynaecol* 1975; **82:** 702–10.

13 Ioannou C, Sarris I, Hoch L, et al. Standardisation of crown-rump length measurement. *BJOG* 2013; **120** (suppl 2)**:** 38–41.

14 Papageorghiou A, Sarris I, Ioannou C, et al. Ultrasound methodology used to construct the fetal growth standards in the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 27–32.

15 Sarris I, Ioannou C, Ohuma E, et al. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 33–37.

16 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.

17 Sarris I, Ioannou C, Dighe M, et al. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound Obstet Gynecol* 2011; **38:** 681–87.

18 Ohuma E, Hoch L, Cosgrove C, et al. Managing data for the international, multicentre INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 64–70.

19 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:** 610–15.

20 Royston P. Constructing time-specific reference ranges. *Stat Med* 1991; **10:** 675–90.

21 Altman D, Ohuma E. Statistical considerations for the development of prescriptive fetal and newborn growth standards in the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 71–76.

22 Altman DG, Chitty LS. Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; **3:** 378–84.

23 Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; **6:** 307–12.

24 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.

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

26   Ioannou C, Talbot K, Ohuma E, et al. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. *BJOG* 2012; **119:** 1425–39.

27   Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat* 1994; **43:** 429–67.

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

29   Cole TJ. Using the LMS method to measure skewness in the NCHS and Dutch National height standards. *Ann Hum Biol* 1989; **16:** 407–19.

30   Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; **11:** 1305–19.

31   Rigby RA, Stasinopoulos DM. Using the Box-Cox *t* distribution in GAMLSS to model skewness and kurtosis. *Stat Model* 2006; **6:** 209–29.

32   Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Stat Med* 2004; **23:** 3053–76.

33   Rasbash J, Charlton C, Browne WJ, Healy M, Cameron B. MLwiN version 2.1. Bristol: Centre for Multilevel Modelling, University of Bristol, 2009.

34   Leckie G, Charlton, C. runmlwin: a program to run the MLwiN multilevel modeling software from within Stata. *J Stat Soft* 2013; **52:** 1–40.

35   Green PJ, Silverman BW. Nonparametric regression and generalized linear models: a roughness penalty approach. London: Chapman & Hall, 1994.

36   Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statist Sci* 1996; **11:** 89–158.

37   Wade A, Kurmanavicius J. Creating unbiased cross-sectional covariate-related reference ranges from serial correlated measurements. *Biostatistics* 2009; **10:** 147–54.

38   Wade AM, Ades AE. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Stat Med* 1998; **17:** 1989–2002.

39   R Development Core Team. R: a language and environment for statistical computing. Austria: R Foundation for Statistical Computing V, 2008.

40   Rigby RA, Stasinopoulos DM. Generalized additive models for Location Scale and Shape (GAMLSS) in R. *J Stat Soft* 2007; **23:** 1–46.

41   Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Appl Statist* 2005; **54:** 507–54.

42   Garza C. New growth standards for the 21st century: a prescriptive approach. *Nutr Rev* 2006; **64:** S55–59.

43   WHO. Physical status: the use and interpretation of anthropometry. *World Health Organ Tech Rep Ser* 1995; **854:** 1–452.

44   Chatfield A, Caglia JM, Dhillon S, et al. Translating research into practice: the introduction of the INTERGROWTH-21st package of clinical standards, tools and guidelines into policies, programmes and services. *BJOG* 2013; **120** (suppl 2): 139–42.

45   Villar J, Cheikh Ismail L, Victora CG, et al, for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014; **384:** 857–68.

46   Salomon LJ, Bernard JP, Duyme M, Buvat I, Ville Y. The impact of choice of reference charts and equations on the assessment of fetal biometry. *Ultrasound Obstet Gynecol* 2005; **25:** 559–65.

47   Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med* 2003; **348:** 1166–70.

48   Kato N. Ethnic differences in genetic predisposition to hypertension. *Hypertension Res* 2012; **35:** 574–81.

49   Papageorghiou AT, Kennedy SH, Salomon LJ, et al. International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown-rump length in the first trimester. *Ultrasound Obstet Gynecol* 2014; published online July 8. DOI:10.1002/uog.13448.

50   Lawn JE, Blencowe H, Oza S, et al, for The Lancet Every Newborn Study Group. Every Newborn: progress, priorities, and potential beyond survival. *Lancet* 2014; **384:** 189–205.