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

Aris T Papageorghiou, Eric O Ohuma, Douglas G Altman, Tullia Todros, Leila Cheikh Ismail, Ann Lambert, Yasin 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-21^st^)

Summary
Background In 2006, WHO produced international growth standards for infants and children up to 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-21^st^ 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 4,607 (35%) were eligible. 4,321 (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: ±2.6 mm (SD 3.0) for abdominal circumference; 0.03 mm (0.8) and –0.05 mm (0.8) for occipitofrontal diameter; 2.99 mm (3.1), 0.25 mm (3.2), and –4.22 mm (3.7) for abdominal circumference; and 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; 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, 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 and the absence of suitable international standards similar to the standards used for monitoring infant growth. 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. The use of such a range of charts and cutoff points in clinical decision making about fetal growth patterns inevitably leads to diagnostic confusion, difficulties comparing outcomes across studies, and the absence of suitable international standards similar to the standards used for monitoring infant growth.

Lancet 2014; 384: 869-79
See Comment page R35
Joint senior authors
†Members listed in the appendix

Young, Abel, Ruyan, Fossey, Paterson, Krestin, for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21^st^)
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. The aim of the Fetal Growth Longitudinal Study (FGLS)—part of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21+) Project—was to construct such standards for fetal growth, adopting the same approach and methods as the WHO Multicentre Growth Reference Study (MGRS) 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.

Methods
Study design and participants
INTERGROWTH-21+ 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. The INTERGROWTH-21+ 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, so as to produce prescriptive growth standards to complement the existing WHO Child Growth Standards, 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-21+ Project, which has been described in detail elsewhere. 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 in those urban areas occurred. For example, adequate nutritional status was defined in criteria and definitions has been published previously. For instance, nutritional status is defined in terms of body mass index (BMI), haemoglobin concentration, and mid-arm circumference.

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. 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.

For the INTERGROWTH-21+ website see www.intergrowth21.org
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-21™ 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-21™ 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). 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. 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. 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. 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. 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-21™ 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. 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.

**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. 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
Articles

included in a table in a previous publication.32 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.33-36 The statistical methods used to construct the fetal standards from this pooled sample were based on the recommendations of Altman and Chitty37 and Royston and Altman,38 complemented by recent scientific literature reviews39-41 and our systematic review of the methods used in previous ultrasound studies aimed at creating references of fetal size.42 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,36 Cole’s lambda (λ), mu (μ), and sigma (σ) (LMS) method,33-34 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 LMST35 (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 LMSP36 (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 models43,44 to account for repeat measurements. Furthermore, to present the curves, we assessed three smoothing techniques: fractional polynomials,36 cubic splines,36 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.45 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;46 therefore, we would expect the fitted centiles to be slightly narrower than empirical ones. All analyses were done in R statistical software47 using the Generalised Additive Models for Location, Scale and Shape (GAMLSS) framework48,49 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-21+ 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
The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. JV, ATP, EOO, FCB, DGA, CGV and SHK had access to the full raw dataset. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results
We screened 13108 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 I). 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

<table>
<thead>
<tr>
<th>Table 1: Baseline characteristics and perinatal events of the Fetal Growth Longitudinal Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pregnancies (n=4321)</strong></td>
</tr>
<tr>
<td>Maternal age, years</td>
</tr>
<tr>
<td>Maternal height, cm</td>
</tr>
<tr>
<td>Maternal weight, kg</td>
</tr>
<tr>
<td>Paternal height, cm</td>
</tr>
<tr>
<td>Body-mass index, kg/m²</td>
</tr>
<tr>
<td>Gestational age at first visit, weeks</td>
</tr>
<tr>
<td>Years of formal education, years</td>
</tr>
<tr>
<td>Haemoglobin level at &lt;15 weeks, g/L</td>
</tr>
<tr>
<td>Married or cohabiting</td>
</tr>
<tr>
<td>Nulliparous</td>
</tr>
<tr>
<td>Pre-eclampsia</td>
</tr>
<tr>
<td>Pyelonephritis</td>
</tr>
<tr>
<td>Any sexually transmitted infection</td>
</tr>
<tr>
<td>Spontaneous initiation of labour</td>
</tr>
<tr>
<td>PPROM (&lt;37 weeks)</td>
</tr>
<tr>
<td>Caesarean section</td>
</tr>
<tr>
<td>NICU admission &gt;1 day</td>
</tr>
<tr>
<td>Preterm (&lt;37 weeks)</td>
</tr>
<tr>
<td>Preterm and spontaneous onset of labour</td>
</tr>
<tr>
<td>Term LBW (&lt;2500 g; ≥37 weeks)</td>
</tr>
<tr>
<td>Neonatal mortality</td>
</tr>
<tr>
<td>Male sex</td>
</tr>
<tr>
<td>Exclusive breastfeeding at discharge</td>
</tr>
<tr>
<td>Mother admitted to intensive care unit</td>
</tr>
<tr>
<td>Birthweight (≥37 weeks), kg</td>
</tr>
<tr>
<td>Birth length (≥37 weeks), cm</td>
</tr>
<tr>
<td>Birth head circumference (≥37 weeks), cm</td>
</tr>
</tbody>
</table>

*Congenital malformations diagnosed by ultrasound during pregnancy or at birth by clinical examination.

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 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.
Discussion
To describe optimal fetal growth for singleton pregnancies, the INTERGROWTH-21+ Project used a prescriptive approach 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 and implemented by the WHO MGRS for infants and children. 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 3.1), compared to 49.4 cm (1-9) in the INTERGROWTH-21+ cohort, and the mean birthweight in the WHO MGRS was 3.3 kg (0.9), 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, have the potential to improve pregnancy outcomes, 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. 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. 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.

Our study aimed to avoid these limitations. 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 of skeletal measures across the study sites. The answer is 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.

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. Additionally, we used a population-based sample of well nourished participants at each of the INTERGROWTH-21 studies—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.

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. We found 337 studies, which had substantial methodological heterogeneity, and none met the WHO recommendations for monitoring fetal anthropometric measurements. 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.

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, 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.
Finally, the cumulative experience gained from implementation in more than 140 countries of the WHO MGRS standards, which were derived from the descriptive, international study of postnatal growth, and whose design and methods were almost identical to the INTERGROWTH-21 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, in addition to our newborn standards, 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”.

Contributors
JV and SHK conceptualised and designed the INTERGROWTH-21 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-21 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.

Acknowledgments
This project was supported by a generous grant from the Bill & Melinda Gates Foundation to the University of Oxford (Oxford, UK), for which we are very grateful. We also thank the Health Authorities in Pelotas, Brazil; Beijing, China; Nagpur, India; Turin, Italy; Nairobi, Kenya; Muscat, Oman; Oxford, UK; and Seattle, WA, USA who facilitated the project by allowing participation of these study sites as coordinating centres. We are extremely grateful to Philips Medical Systems who provided the ultrasound equipment and technical assistance throughout the project. We also thank MedSciNet UK for setting up the INTERGROWTH-21 website and for the development, maintenance, and support of the online data management system. We thank the parents and infants who participated in the studies and the more than 200 members of the research teams who made the implementation of this project possible. The participating hospitals included: Brazil, Pelotas (Hospital Miguel Pichet, Hospital São Francisco de Paula, Santa Casa de Misericórdia de Pelotas, and Hospital Escola da Universidade Federal de Pelotas); China, Beijing (Beijing Obstetrics and Gynecology Hospital, Shunyi Maternal and Child Health Centre, and Shunyi General Hospital); India, Nagpur (Ketkar Hospital, Avanti Institute of Cardiology Private Limited, Avanti Hospital, Gurukrupa Maternity Hospital, Mulik Hospital and Research Centre, Nandikö Hospital, Om Women’s Hospital, Renuka Hospital and Maternity Home, Saboo Hospital, Brahmachari Taori Memorial Hospital, and Somani Nursing Home); Kenya, Nairobi (Aga Khan University Hospital, MP Shah Hospital, and Avenue Hospital); Italy, Turin (Ospedale Infantile Regina Margherita Sant’ Anna and Azienda Ospedaliera Ordine Mauriziano); Oman, Muscat (Khoulia Hospital, Royal Hospital, Wattaya Health Centre, and Wattaya Obstetrics and Gynaecology Poly Clinic, Wattaya Health Centre, Rusvi Health Centre, Al-Ghoubra Health Centre, and Al-Khawair Health Centre); UK, Oxford (John Radcliffe Hospital and USA, Seattle (University of Washington Hospital, Swedish Hospital, and Providence Everett Hospital). Full acknowledgment to all those who contributed to the development of the INTERGROWTH-21 Project protocol is presented on the INTERGROWTH-21 website.

References


International fetal growth standards: one size fits all

Those working with pregnant women know the situation well: whether a particular unborn baby is judged to be too small will partly depend on the choice of fetal biometric charts used. Some charts can increase the likelihood of finding a fetal measurement below a threshold for smallness by fourfold or more. Inconsistent chart use and overestimation of fetal smallness can result in cynicism, confusion, and anxiety for pregnant women and their caregivers at all stages of pregnancy. The stakes are especially high when smallness of biometric measures alone raises a red flag for aneuploidy, other genetic syndromes such as skeletal dysplasia, or fetal developmental abnormalities such as microcephaly. In later pregnancy, suspected fetal smallness might trigger interventions such as induction of labour and caesarean section, with consequent risks, including iatrogenic prematurity. Human, pregnancy, and financial costs are likely to follow.

However, to miss true fetal growth restriction is equally serious because suboptimum birthweight is associated with lifelong health disadvantages. Most pertinent to maternity caregivers is that fetal growth restriction is the single biggest risk factor for stillbirth. The fourfold increase in stillbirth risk in the growth-restricted fetus is doubled again in those in whom fetal growth restriction is undetected. An accepted universal standard of fetal size and growth is urgently needed.

Robust methods for fetal biometric chart development have been described since the 1990s, but many frequently used charts do not adhere to these standards. Practitioners might not be able to identify and access a methodologically superior chart easily, and instead default to what is loaded onto their ultrasound machine or follow institutional practices, which—clinicians acknowledge—can vary according to where the practitioner is working on different days of the week.

The multicentre, population-based Fetal Longitudinal Growth Study (FLGS) of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21) Project cuts through the noise with a clear signal. Led by researchers from around the world, this study has followed the methods of the WHO Multicentre Growth Reference Study, recruiting more than 4000 pregnant women from eight countries. Although differing in ethnicity, these women were intentionally sampled for their shared trait of being at low risk of pregnancy complications attributable to malnutrition, obesity, socioeconomic deprivation, or major environmental pollution. A recent publication from this group came to the conclusion that fetuses and newborn infants grow very similarly in these eight different research settings.

This information is especially valuable for the multicultural societies that now exist in many European,
North American, and Australasian countries, which are characterised by high rates of immigration and intermarriage. To expect all fetuses to have the same growth potential irrespective of ethnic origin is simple and fair: identification of ethnic origin might not be straightforward in multicultural societies, and to adapt fetal growth expectations for ethnic origin creates the risk of generating an expectation of poorer perinatal outcome.22 The latter can detract from efforts to improve perinatal outcomes for vulnerable groups. Arguably, the most important conclusion that can be drawn from the INTERGROWTH-21st Consortium’s findings is that recorded intercountry differences in fetal size and growth are more indicative of deprivation than geographical location. Acceptance of international growth standards involves a change in theoretical approach and sets the bar for universal aspirational targets of fetal growth.

In The Lancet, Aris Papageorghiou and colleagues21 for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) report FLGS charts for fetal biparietal diameter, head circumference, occipitofrontal diameter, abdominal circumference, and femur length, which were obtained every 5 weeks (within 1 week either side) from 14 weeks to 42 weeks of gestation. These charts have been developed with great care, including sonographer training, quality assurance, attention to minimisation of bias, data management, and statistical and ethical aspects, which have been described previously.23 No issue was too small for consideration. For example, questions such as whether the biparietal diameter should be measured outer-to-inner—as in the new FLGS charts—or outer-to-inner were discussed with guideline writers, opinion leaders, and experienced sonographers, with the final decision favouring continuity with postnatal measurements.31

Of particular importance, Papageorghiou and colleagues achieved a sample size of 4321, which is sufficiently large to produce reliable extreme centiles that are of clinical significance. The charts were not artificially modified by post-hoc truncation to produce a supernal middle range of measurements such as 10th–90th centiles, but present aspirational standards with valid extreme measurement values. The 3rd, 5th, 10th, 50th, 90th, 95th, and 97th centile curves were calculated according to gestational age for each of the five ultrasound measures, representing the international standards for fetal growth. The mean differences between the observed and smoothed centiles for the clinically important 3rd and 97th centiles, respectively, were small: 2·25 mm and –2·69 mm for head circumference; 0·83 mm and –0·84 mm for biparietal diameter; 0·63 mm and –1·05 mm for occipitofrontal diameter; 2·99 mm and –4·22 mm for abdominal circumference; and 0·62 mm and –0·65 mm for femur length. To put these differences into context, they are five-to-ten-times smaller than the differences between varying charts.

We strongly encourage fetal medicine, obstetrics, and midwifery professional bodies to adopt and endorse consistent use of these high-quality, globally relevant charts, which give the best available answer to how unborn babies should be growing.

*Elizabeth A McCarthy, Susan P Walker
Department of Obstetrics and Gynaecology, Mercy Hospital for Women, University of Melbourne, Heidelberg 3084, VIC, Australia eamcca@unimelb.edu.au

We declare no competing interests.