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

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

Summary

Background Large differences exist in size at birth and in rates of impaired fetal growth worldwide. The relative effects of nutrition, disease, the environment, and genetics on these differences are often debated. In clinical practice, various references are often used to assess fetal growth and newborn size across populations and ethnic origins, whereas international standards for assessing growth in infants and children have been established. In the INTERGROWTH-21st Project, our aim was to assess fetal growth and newborn size in eight geographically defined urban populations in which the health and nutrition needs of mothers were met and adequate antenatal care was provided.

Methods For this study, fetal growth and newborn size were measured in two INTERGROWTH-21st component studies using prespecified markers and the same methods, equipment, and selection criteria. In the Fetal Growth Longitudinal Study (FGLS), we studied educated, affluent, healthy women, with adequate nutritional status who were at low risk of intrauterine growth restriction. The primary markers of fetal growth were ultrasound measurements of fetal crown-rump length at less than 14 weeks and 0 days of gestation and fetal head circumference from 14 weeks and 0 days to 40 weeks and 0 days of gestation, and birthlength for newborn size. In the concomitant, population-based Newborn Cross-Sectional Study (NCSS), we measured birthlength in all newborn babies from the eight geographically defined urban populations with the same methods, instruments, and staff as in FGLS. From this large NCSS cohort, we selected an FGLS-like subpopulation to match FGLS with the same eligibility criteria.

Findings Between May 14, 2009, and Aug 2, 2013, we enrolled 4607 women in FGLS and 59137 women in NCSS. From NCSS, 20486 (34·6%) women met the FGLS eligibility criteria, and constituted the FGLS-like subpopulation. With variance component analysis, only between 1·9% and 3·5% of the total variability in crown-rump length, fetal head circumference, and newborn birthlength could be attributed to between-site differences. With standardised site effect analysis in 16 gestational age windows from 9 weeks and 0 days of gestation to birth for the three measures (128 comparisons), only one was marginally higher than 0·5 SD of the standardised site difference range. Sensitivity analyses, excluding individual populations in turn from the pooling of all-site centiles across gestational ages, showed no noticeable effect on the 3rd, 50th, and 97th centiles derived from the remaining populations. Our populations were consistent at birth with those in the WHO Multicentre Growth Reference Study (MGRS). The mean birthlength for term newborn babies in that study was 49·5 cm (SD 1·9), which was very similar to that in the FGLS cohort (49·4 cm [1·9]) and the NCSS derived FGLS-like subpopulation (49·3 cm [1·8]).

Interpretation Fetal growth and newborn length are similar across diverse geographical settings when mothers' nutritional and health needs are met, and environmental constraints on growth are low. The findings for birthlength are in strong agreement with those of the WHO MGRS. These results provide the conceptual frame to create international standards for growth from conception to newborn baby, which will extend the present infant to childhood WHO MGRS standards.

Funding Bill & Melinda Gates Foundation.

Lancet Diabetes Endocrinol 2014; 2: 781–92

Published Online
July 7, 2014
http://dx.doi.org/10.1016/S2213-8587(14)70121-4

See Comment page 766

*Joint senior authors

Other members and committees of INTERGROWTH-21st are listed in the appendix

Nuffield Department of Obstetrics & Gynaecology and Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK (Prof J Villar MD, A T Papageorghiou MD, E O Ohuma MSc, L Cheikh Ismail PhD, A Lambert PhD, Prof S H Kennedy MD); **School of Public Health, Peking University, Beijing, China** (Prof R Pang MD); **Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Oxford, UK** (E O Ohuma, Prof D G Altman DSc); **Programa de Pós-Graduação em Saúde e Comportamento, Universidade Católica de Pelotas, Pelotas, RS, Brazil** (Prof F C Barros MD); **Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil** (Prof C G Victora MD); **Faculty of Health Sciences, Aga Khan University, Nairobi, Kenya** (M Carvalho MD); **Department of Family and Community Health, Ministry of Health, Muscat, Sultanate of Oman** (Y A Jaffer MD); **Dipartimento di Scienze Pediatriche e dell'Adolescenza, Cattedra di Neonatologia, Università degli Studi di Torino, Torino, Italy** (Prof E Bertino MD); **University of Washington School of**

Introduction

Many populations are exposed to adverse environmental conditions and inadequate nutritional intakes that affect fetal growth.[1] Therefore, findings of an increased number of newborn babies small for gestational age in these geographical areas and in immigrants in ethnically heterogeneous populations in developed countries (eg, Netherlands[2] and the USA[3]) are not surprising. However, investigators have attributed the high rates of small for gestational age newborn babies reported in certain populations to genetic factors,[4] despite findings from epidemiological and clinical studies that have consistently shown similar growth patterns across some ethnic groups in infants and

Medicine, Seattle, WA, USA (M G Gravett MD); **Nagpur INTERGROWTH-21st Research Centre, Ketkar Hospital, Nagpur, India** (M Purwar MD); **Center for Perinatal Studies, Swedish Medical Center, Seattle, WA, USA** (I O Frederick PhD); **Department of Engineering Science, University of Oxford, Oxford, UK** (Prof J A Noble DPhil); **Center of Excellence in Women and Child Health, The Aga Khan University, Karachi, Pakistan** (Prof Z A Bhutta PhD); **and Center for Global Child Health, Hospital for Sick Children, Toronto, Canada** (Prof Z A Bhutta)

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

See Online for appendix

children from affluent, well-nourished and geographically diverse backgrounds.[5,6]

Up to now, the strongest scientific evidence supporting the notion, first proposed by Habicht and colleagues,[7] that both infant and child growth are more affected by health, socioeconomic status, and environmental conditions than by ethnic differences, has been provided by the multiethnic WHO Multicentre Growth Reference Study (MGRS) of healthy, breastfed children with minimum environmental, health, and nutrition constraints on growth from six populations in Brazil, Ghana, India, Norway, Oman, and the USA (n=8406).[8,9] Results of the study showed striking similarity in linear growth in children from the six sites,[10] thereby justifying pooling data to construct one international growth standard from birth to 5 years of age, which has since been adopted worldwide.[11,12]

Although ample data have contributed to devising international growth standards for infants and children, so far, the data for fetal growth and newborn size have been limited. The conclusions of two recent systematic reviews[13,14] strongly support the need to develop international standards to assess growth patterns in the prenatal and neonatal periods. Therefore, our aim was to assess fetal growth and newborn size across different populations by mapping skeletal growth as a continuous process from after conception to birth in a prospective, population-based project. We used identical methods in eight geographically diverse urban areas in which mothers' health and nutritional needs were met; sanitation practices and the environment were judged not to be constraining growth; and adequate, standardised antenatal care was provided. If the data generated were consistent with the WHO MGRS standards (birth to 5 years), a global set of international fetal and newborn standards could be generated to allow growth to be monitored from the post-conception period to childhood.

Methods

Study design and participants

For the **protocol** see http://www.intergrowth21.org.uk

INTERGROWTH-21st was a multicentre, multi-ethnic, population-based project, done between April 27, 2009, and March 2, 2014, in eight study sites: the cities of Pelotas (Brazil), Turin (Italy), Muscat (Oman), Oxford (UK), and Seattle (USA); Shunyi County, Beijing (China); the central area of Nagpur (India); and the Parklands suburb of Nairobi (Kenya).[15] Its main aim was to study growth, health, nutrition, and neurodevelopment from less than 14 weeks of gestation to 2 years of age, with the same conceptual framework as the WHO MGRS.[9]

The methods for the project's two component studies presented here (the Fetal Growth Longitudinal Study [FGLS] and the Newborn Cross-Sectional Study [NCSS]) have been described elsewhere in detail.[15] Briefly, the populations that contributed participants to the project were first selected at the cluster level and then at the individual level within each study site. At the cluster level, an urban area (eg, a complete city, county, or part of a city with clear political or geographical limits) was chosen where most deliveries happened in health institutions. The areas had to be located at an altitude of 1600 m or lower. Women had to plan to deliver in these institutions or in a similar hospital located in the same geographical area. Major, known, non-microbiological contamination such as pollution, domestic smoke, radiation, or any other toxic substances, had to be absent or at low levels, as assessed with a data collection form specifically developed for the project.[16]

In the eight urban areas, we selected all institutions providing pregnancy and intrapartum care where more than 80% of deliveries for the target population took place. At the site in central Nagpur, we identified specifically all institutions classified locally as private or corporation hospitals, or those serving the middle to upper socioeconomic population because all institutional deliveries from the target population took place there. From these institutions, we selected the ten largest covering more than 80% of deliveries in the central Nagpur area.

At the individual level, we recruited mothers (and their newborn babies) for FGLS aged older than or equal to 18 years and younger than or equal to 35 years, who measured greater than or equal to 153 cm in height, had BMI greater than or equal to 18·5 kg/m² and less than 30 kg/m², who had no clinically relevant obstetric or gynaecological history, initiated antenatal care at less than 14 weeks of gestation (by menstrual dates), and met the entry criteria of optimum health, nutrition, education, and socioeconomic status.[15] Eligible women were recruited consecutively at each antenatal clinic up to a weekly limit (roughly six women per week) to avoid overwhelming the capacity of the project's ultrasound research team.

In the population-based component (NCSS), we aimed to include concomitantly all babies born in the same eight urban areas that delivered in the institutions chosen (the same institutions as in FGLS plus additional institutions) during 12 months or until the target sample of 7000 newborn babies per site was attained with the same data collection forms, manuals, instruments, and research staff as in FGLS. For analysis purposes, all the NCSS pregnancies were divided into two groups. The first consisted of all pregnancies and newborn babies from women who met all the FGLS demographic, clinical, social and educational entry criteria[15] so as to compare newborn size across similarly defined low-risk populations. This group was named the FGLS-like subpopulation. As expected, not all women in this subpopulation could be included in FGLS (even if they were potentially eligible) because FGLS required a smaller sample and had a limited number of ultrasound scans that could be done per week under the study's carefully controlled conditions. The second group, composed of the remaining newborn babies from higher risk pregnancies, is not considered further here, but will

be described in subsequent publications. The total population of the INTERGROWTH-21st Project consisted of 59 137 women from NCSS of whom 20 486 (34·6%) were considered FGLS-like. 20 486 women met the FGLS selection criteria but only 4607 were recruited at less than 14 weeks of gestation prospectively to FGLS and had scans every 5 weeks. The remaining 15 879 were enrolled at birth.

Procedures

In FGLS, we used the last menstrual period to calculate gestational age provided that: the 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 last menstrual period and crown-rump length, measured by ultrasound at 9 weeks and 0 days to 13 weeks and 6 days from the last menstrual period with a recognised chart,[17] was 7 days or less. The crown-rump length technique was standardised across sites and all ultrasonographers were trained uniformly with strict quality control measures.[18]

All participating hospitals agreed to a policy of routine estimation of gestational age by ultrasound, allowing the eligibility criteria from FGLS to be adopted in the NCSS FGLS-like subpopulation. Gestational age for the NCSS population was estimated with crown-rump length less than 14 weeks and 0 days of gestation or biparietal diameter when antenatal care started between 14 weeks and 0 days and 24 weeks and 0 days of gestation. If the ultrasound estimation was made at more than 24 weeks of gestation, the measurement was only accepted as reliable if any discrepancy between this gestational age and the one based on last menstrual period was 7 days or less.[19]

In FGLS, the protocol required all women to have scans every 5 weeks (within 1 week either side) from the initial dating scan—ie, possible ranges after the dating scan were 14–18, 19–23, 24–28, 29–33, 34–38, and 39–42 weeks of gestation. The ultrasound machine used at all sites (Philips HD-9, Philips Ultrasound, USA, with curvilinear abdominal transducers C5-2, C6-3, V7-3) was specially adapted to allow masking—ie, the measurements were not visible on the screen.[20] At each visit, a set of standard ultrasound measurements were obtained three times from three separately obtained images of each structure. The fetal measurements were biparietal diameter, head circumference, abdominal circumference, and femur length. All ultrasound data were submitted electronically.

After each set of measurements (submitted before scoring), ultrasonographers scored the quality of their images on the basis of standard image-scoring criteria.[21] Ultrasonographers were asked to repeat images that did not achieve the maximum score until they were satisfied the best possible image had been achieved. The training, standardisation and quality control methods used across all sites are described in detail elsewhere.[21] Only after three measurements of each structure were taken, were the averages revealed for clinical purposes.

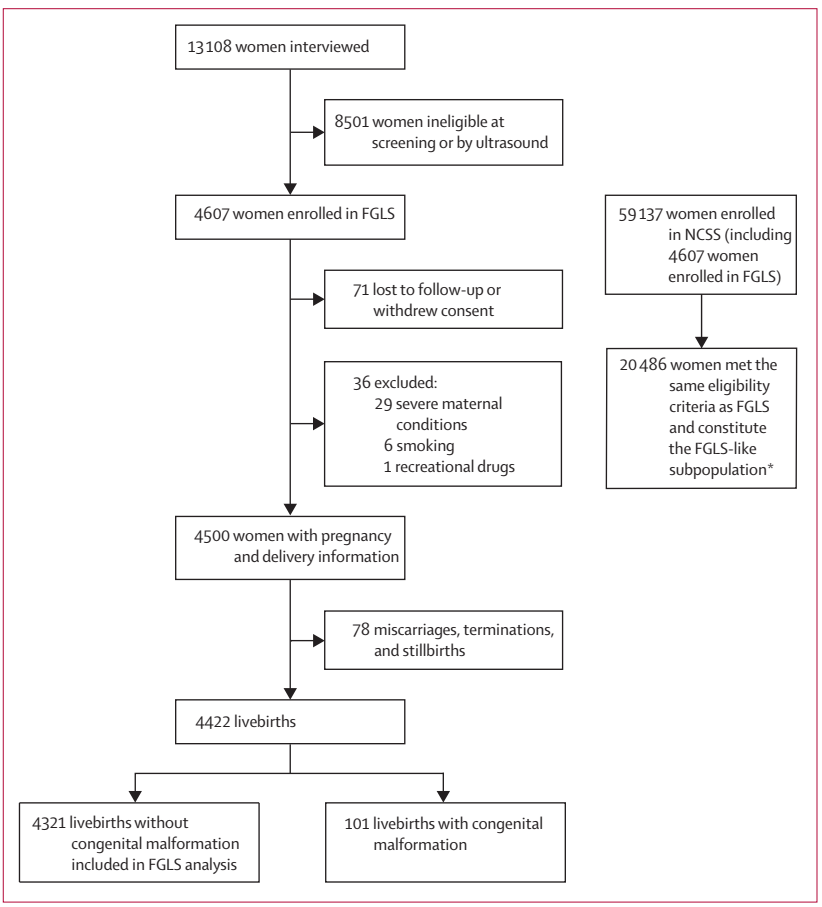

Figure 1: **FGLS and NCSS study profiles**
Congenital malformations were diagnosed by ultrasound during pregnancy or at birth by clinical assessment. FGLS=Fetal Growth Longitudinal Study. NCSS=Newborn Cross-Sectional Study. *Data used in this analysis were collected for all women at the same time with the same data collection instruments and methods.

We obtained newborn anthropometric measures in both FGLS and NCSS within 12 h of birth with identical equipment at all sites: electronic scale for birthweight (Seca, Hangzhou, China), recumbent birthlength with a specially designed Harpenden infantometer (Chasmors, London, UK), and head circumference with a metallic non-extendable tape (Chasmors).[22] The equipment, which was calibrated two times a week, was selected for accuracy, precision, and robustness as shown in previous studies.[23] Measurement procedures were standardised on the basis of WHO recommendations.[23] The intra-observer and inter-observer error of measurement values for head circumference ranged from 0·3 to 0·4 cm, and for recumbent length from 0·3 to 0·5 cm during the standardisation sessions. Each measurement was collected independently by two anthropometrists.[24] If the difference between their measurements exceeded the set maximum allowable difference (birthweight 50 g; birthlength 7 mm and head circumference 5 mm), both observers had independently to retake that

	Brazil		China		India		Italy		Kenya		Oman		UK		USA		Total	
	FGLS (n=411)	NCSS FGLS-like sub-population (n=1595)	FGLS (n=609)	NCSS FGLS-like sub-population (n=3551)	FGLS (n=625)	NCSS FGLS-like sub-population (n=2493)	FGLS (n=509)	NCSS FGLS-like sub-population (n=2358)	FGLS (n=617)	NCSS FGLS-like sub-population (n=3702)	FGLS (n=599)	NCSS FGLS-like sub-population (n=2821)	FGLS (n=640)	NCSS FGLS-like sub-population (n=2939)	FGLS (n=311)	NCSS FGLS-like sub-population (n=1027)	FGLS (N=4321)	NCSS FGLS-like sub-population (N=20486)
Maternal age (years)	28·3 (4·1)	26·4 (4·8)	26·8 (2·9)	26·3 (3·0)	27·3 (3·5)	27·5 (3·3)	29·0 (4·3)	29·9 (4·0)	28·9 (3·5)	28·8 (3·5)	27·2 (3·8)	26·9 (4·0)	30·2 (3·7)	29·1 (4·3)	29·8 (3·6)	29·5 (3·9)	28·4 (3·9)	28·0 (4·0)
Maternal height (cm)	163·3 (5·4)	162·5 (5·4)	161·5 (4·7)	161·7 (4·5)	158·3 (3·6)	157·6 (3·3)	162·8 (5·8)	163·3 (5·6)	162·1 (5·5)	162·3 (5·5)	159·7 (4·3)	158·8 (4·1)	166·2 (5·9)	165·3 (6·1)	166·0 (6·5)	164·8 (6·2)	162·2 (5·8)	161·8 (5·6)
Maternal weight (kg)	64·5 (8·5)	63·2 (8·4)	57·9 (7·6)	58·8 (7·6)	56·1 (8·1)	57·0 (7·7)	61·1 (8·1)	60·4 (7·9)	63·1 (8·9)	63·6 (8·5)	60·6 (8·9)	60·7 (8·5)	64·4 (8·9)	64·4 (8·8)	65·7 (9·9)	63·7 (9·0)	61·3 (9·1)	61·3 (8·6)
Paternal height (cm)	174·1 (6·2)	..	174·5 (4·8)	..	168·5 (5·7)	..	176·7 (7·6)	..	175·0 (6·0)	..	171·9 (6·2)	..	179·5 (6·8)	..	180·6 (7·6)	..	174·4 (7·3)	..
BMI (kg/m^2)	24·2 (2·8)	23·9 (2·8)	22·2 (2·6)	22·5 (2·7)	22·4 (3·0)	22·9 (2·9)	23·0 (2·8)	22·6 (2·6)	24·0 (3·0)	24·1 (2·9)	23·7 (3·2)	24·1 (3·1)	23·3 (2·8)	23·5 (2·8)	23·8 (3·1)	23·4 (2·8)	23·3 (3·0)	23·4 (2·9)
Gestational age at first visit (weeks)	11·5 (1·4)	14·0 (5·8)	12·8 (0·9)	16·2 (5·4)	11·6 (1·5)	14·3 (7·5)	11·9 (1·0)	13·1 (3·6)	11·5 (1·5)	17·1 (7·9)	10·9 (1·4)	15·2 (5·7)	12·3 (1·1)	13·2 (3·1)	11·6 (1·8)	12·0 (4·0)	11·8 (1·4)	14·8 (6·0)
Years of formal education	13·6 (3·3)	11·3 (3·6)	14·2 (2·2)	13·9 (1·9)	16·1 (1·3)	16·2 (1·3)	13·5 (4·0)	13·7 (3·8)	15·5 (1·4)	14·9 (2·3)	14·3 (2·5)	13·2 (2·8)	16·0 (3·0)	16·0 (3·0)	17·1 (2·8)	16·5 (3·2)	15·0 (2·8)	14·2 (3·0)
Haemoglobin concentration <15 weeks (g/dL)	12·4 (0·9)	12·3 (0·9)	13·5 (0·9)	13·3 (1·0)	11·6 (0·6)	11·2 (1·1)	12·9 (0·9)	12·9 (1·0)	12·9 (1·3)	12·5 (1·4)	11·7 (1·1)	11·7 (1·1)	12·7 (0·9)	12·5 (0·9)	12·7 (0·9)	12·6 (0·9)	12·5 (1·1)	12·3 (1·2)
Married or co-habiting	393 (95·6%)	1468 (92·0%)	607 (99·7%)	3548 (99·9%)	619 (99·0%)	2485 (99·7%)	493 (96·9%)	2327 (98·7%)	558 (90·4%)	3525 (95·2%)	599 (100·0%)	2821 (100·0%)	632 (98·8%)	2762 (94·0%)	303 (97·4%)	941 (91·6%)	4204 (97·3%)	19877 (97·0%)
Nulliparous	301 (73·2%)	998 (62·6%)	588 (96·6%)	3320 (93·5%)	443 (70·9%)	1719 (69·0%)	328 (64·4%)	1472 (62·4%)	385 (62·4%)	1877 (50·7%)	328 (54·8%)	1228 (43·5%)	378 (59·1%)	1753 (59·6%)	204 (65·6%)	629 (61·2%)	2955 (68·4%)	12996 (63·4%)

Data are n (SD) or n (%). Only pregnancies with a live singleton birth and no congenital malformation included. FGLS=Fetal Growth Longitudinal Study. NCSS=Newborn Cross-Sectional Study.

Table 1: **Baseline characteristics for women enrolled in the FGLS and the FGLS-like subpopulation from the NCSS**

measurement a second and, if necessary, a third time. The training, standardisation, and quality control methods employed across all sites are described elsewhere.[24]

Neonatal clinical practices, including neonatal intensive care unit care and feeding, were also standardised across sites according to a basic package of internationally accepted evidence-based practices, following an agreed protocol adopted by the project's Neonatal Study Group.[25]

We decided a priori[15] to assess the similarities between fetal growth and newborn size with the following fat-free mass indicators: crown-rump length less than 14 weeks and 0 days of gestation, fetal head circumference from 14 weeks and 0 days of gestation onwards, and birthlength.[10] We also obtained, as a complementary measurement, head circumference at birth to match fetal head circumference because it is the only skeletal measure available from early pregnancy to childhood. Therefore, we obtained information about early fetal size, longitudinal fetal growth, and newborn size for the complete FGLS cohort, and newborn size for the NCSS FGLS-like subpopulation in the eight study sites, with the same standardised procedures.

All documentation used in the INTERGROWTH-21st studies was tested locally and introduced into the specially developed, online electronic data entry, cleaning, and management system MedSciNet. The average values of the repeated ultrasound and anthropometric measures were used in the present analyses.[26]

Statistical analysis

The sample size for FGLS was based on pragmatic and statistical considerations. Statistical considerations focused on the precision and accuracy of a single centile and regression-based reference limits.[27,28] We established an average target sample of 500 pregnant women per study site, after exclusion of complicated pregnancies and those lost to follow-up.[29] This sample size was regarded as adequate to explore site-specific differences or subgroups of 10% of the whole study if needed.

Overall, we expected roughly 3% of patients would be lost to follow-up and that another 3% would be excluded (with prespecified criteria) from the study population because of fetal deaths and congenital abnormalities. We also excluded mothers diagnosed with catastrophic or very severe medical conditions (eg, cancer or HIV); those with severe unanticipated pregnancy-related conditions requiring 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 a malaria episode).

Fetal measurements tend generally to be close to a normal distribution at each specific gestational age[30] and are, therefore, presented as the mean and SD. We used three complementary analytic strategies: firstly, we used analysis of variance (variance component analysis) to calculate the percentage of variance in the

cross-sectional measures (crown-rump length and birthlength) due to between-sites variance.[10] Because fetal head circumference measurements were taken several times during pregnancy from the same fetuses, we also estimated variance in individuals within a site (within-site variance). We applied a multilevel, mixed effect, linear regression analysis for cross-sectional and repeated measures as appropriate, adjusting for gestational age. We treated gestational age as a fixed effect, whereas sites and individuals were treated as random effects.

Secondly, for each site and for each measure, at 16 specified gestational age windows (ten fetal and six newborn), we calculated the difference between the mean from that site and the mean of all sites together. Each difference was then expressed as a proportion of all the sites' SD (ie, SD of the data pooled across all sites) at each corresponding gestational age to give the standardised site difference,[10] which is similar to a *Z* score and is expressed in units of the all sites' SD (ie, 1·0 standardised site difference=1·0 all sites' SD). The standardised site difference allows for direct comparisons of different biometric measures in populations across the antenatal (crown-rump length and fetal head circumference) and newborn (birthlength and head circumference at birth) periods, all standardised by the corresponding pooled SD. A pattern of standardised site difference values of less than 0·5 was prespecified in the project protocol as adequate for the combination of data from all sites following the cutoff point used in the study to create international standards for infant and child growth (WHO MGRS).[10]

Because some variability existed across sites in the gestational age distribution within each gestational age window at which fetal head circumference data were obtained, we adjusted the fetal head circumference measures at each site to the midpoint of each gestational age interval by use of estimates obtained from fitting a fractional polynomial regression model, assuming the growth rate was uniform within each of the 5 week intervals.

Thirdly, we did a sensitivity analysis to assess the effect of exclusion of data from single sites. We compared standardised site differences for the full dataset and the reduced datasets (one site excluded at a time). Additionally, we compared the 3rd, 50th, and 97th centiles on the basis of fractional polynomial regression between the complete dataset and the reduced datasets (one site excluded at a time). For each measure we compared the graphs visually, supported by calculations of standardised site difference for each centile within 16 gestational age windows. We did all analyses with STATA (version 11.2).

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. Women in FGLS gave written consent; in NCSS, we obtained institutional consent to use routinely collected data and women gave oral consent.

Role of the funding source

The funders had no role in the study design, data collection, analysis, interpretation of the data, or writing of the report. The following authors had access to the full raw dataset: JV, EOO, FCB, DGA, CGV, and SHK. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Between April 27, 2009, and Aug 2, 2013, in FGLS, we screened 13 108 pregnant women attending the study clinics; of these, 4607 (35%) who met the eligibility criteria[15] consented and were enrolled (figure 1). The most common reasons for ineligibility were maternal age younger than 18 years or older than 35 years (915, 11%), maternal height less than 153 cm (1022, 12%; mostly in India and Oman), and BMI of 30 kg/m^2 or higher (1009, 12%; mostly in the UK and USA). The contribution of each site to the total study population ranged from 7% (311/4607) in the USA to 14% (640/4607) in the UK. Of the 4607 enrolled, we excluded 36 women (0·8%) who developed severe conditions during pregnancy or took up smoking or drug use, and 71 (1·5%) were lost to follow-up or withdrew consent. Of the 4422 women who had live singleton births, 4321 (98%) had newborn babies without congenital malformations; their data comprised the FGLS population in this study.

For **MedSciNet** see http://medscinet.com

In NCSS, 59 137 women were considered for enrolment. Of these, 20 486 (35%) met the same eligibility criteria as FGLS,[15] had a reliable estimate of gestational age from ultrasound, and delivered a live singleton newborn without a congenital malformation. These 20 486 women formed the FGLS-like subpopulation. The most common reasons for ineligibility for the FGLS-like subpopulation (and some women had more than one factor) were maternal age younger than 18 years or older than 35 years (7929 of 31 685 ineligible women with singleton births and reliable dating, 25%), maternal height less than 153 cm (5932, 19%), BMI 30 kg/m^2 or higher (6579, 21%), and previous obstetric history (8406, 26%).

The contribution by site to the NCSS FGLS-like subpopulation ranged from 5% in the USA and 8% in Brazil to 18% in Kenya. Differences in each country's contribution to the sample size for the FGLS-like subpopulation were indicative of risk profiles of the underlying populations in the selected urban populations—ie, inner city Seattle and Pelotas had fewer eligible women than did the other study areas.

At entry, the FGLS populations in the eight sites were, as expected, similar because the same entry criteria were used (table 1). We noted differences in maternal and

	Brazil		China		India		Italy		Kenya		Oman		UK		USA		Total	
	FGLS (n=411)	NCSS FGLS-like sub-population (n=1595)	FGLS (n=609)	NCSS FGLS-like sub-population (n=3551)	FGLS (n=625)	NCSS FGLS-like sub-population (n=2493)	FGLS (n=509)	NCSS FGLS-like sub-population (n=2358)	FGLS (n=617)	NCSS FGLS-like sub-population (n=3702)	FGLS (n=599)	NCSS FGLS-like sub-population (n=2821)	FGLS (n=640)	NCSS FGLS-like sub-population (n=2939)	FGLS (n=311)	NCSS FGLS-like sub-population (n=1027)	FGLS (N=4321)	NCSS FGLS-like sub-population (N=20 486)
Pre-eclampsia	0 (0·0%)	23 (1·4%)	8 (1·3%)	49 (1·4%)	2 (0·3%)	6 (0·2%)	2 (0·4%)	13 (0·6%)	6 (1·0%)	40 (1·1%)	1 (0·2%)	8 (0·3%)	7 (1·1%)	102 (3·5%)	5 (1·6%)	15 (1·5%)	31 (0·7%)	256 (1·2%)
Pyelonephritis	3 (0·7%)	25 (1·6%)	0 (0·0%)	0 (0·0%)	0 (0·0%)	0 (0·0%)	3 (0·6%)	4 (0·2%)	6 (1·0%)	16 (0·4%)	2 (0·3%)	3 (0·1%)	1 (0·2%)	2 (0·1%)	1 (0·3%)	4 (0·4%)	16 (0·4%)	54 (0·3%)
Any sexually transmitted infection	0 (0·0%)	20 (1·3%)	0 (0·0%)	0 (0·0%)	0 (0·0%)	0 (0·0%)	1 (0·2%)	8 (0·3%)	0 (0·0%)	2 (0·1%)	0 (0·0%)	0 (0·0%)	1 (0·2%)	2 (0·1%)	1 (0·3%)	36 (3·5%)	3 (0·1%)	68 (0·3%)
Spontaneous initiation of labour	138 (33·6%)	850 (53·3%)	298 (48·9%)	1390 (39·1%)	421 (67·4%)	1528 (61·3%)	395 (77·6%)	1985 (84·2%)	430 (69·7%)	2482 (67·0%)	530 (88·5%)	2494 (88·4%)	450 (70·3%)	2025 (68·9%)	206 (66·2%)	716 (69·7%)	2868 (66·4%)	13 470 (65·8%)
PPROM (<37^{+0} weeks)	19 (4·6%)	62 (3·9%)	8 (1·3%)	65 (1·8%)	9 (1·4%)	48 (1·9%)	10 (2·0%)	24 (1·0%)	12 (1·9%)	47 (1·3%)	10 (1·7%)	35 (1·2%)	6 (0·9%)	37 (1·3%)	6 (1·9%)	20 (1·9%)	80 (1·9%)	338 (1·6%)
Caesarean section	341 (83·0%)	1040 (65·2%)	321 (52·7%)	2077 (58·5%)	326 (52·2%)	1516 (60·8%)	98 (19·3%)	488 (20·7%)	192 (31·1%)	1187 (32·1%)	83 (13·9%)	395 (14·0%)	101 (15·8%)	513 (17·5%)	79 (25·4%)	236 (23·0%)	1541 (35·7%)	7452 (36·4%)
NICU admission >1 day	30 (7·3%)	143 (9·0%)	56 (9·2%)	438 (12·3%)	27 (4·3%)	93 (3·7%)	17 (3·3%)	56 (2·4%)	25 (4·1%)	143 (3·9%)	44 (7·3%)	152 (5·4%)	24 (3·8%)	108 (3·7%)	17 (5·5%)	51 (5·0%)	240 (5·6%)	1184 (5·8%)
Preterm (<37^{+0} weeks)	40 (9·7%)	143 (9·0%)	16 (2·6%)	212 (6·0%)	31 (5·0%)	250 (10·0%)	21 (4·1%)	83 (3·5%)	31 (5·0%)	154 (4·2%)	24 (4·0%)	145 (5·1%)	21 (3·3%)	100 (3·4%)	11 (3·5%)	49 (4·8%)	195 (4·5%)	1136 (5·5%)
Preterm and spontaneous onset of labour	20 (4·9%)	79 (5·0%)	11 (1·8%)	87 (2·5%)	18 (2·9%)	111 (4·5%)	18 (3·5%)	55 (2·3%)	19 (3·1%)	91 (2·5%)	18 (3·0%)	113 (4·0%)	13 (2·0%)	57 (1·9%)	9 (2·9%)	41 (4·0%)	126 (2·9%)	634 (3·1%)
Term low birthweight (<2500 g; ≥37^{+0} weeks)	4 (1·0%)	31 (1·9%)	4 (0·7%)	22 (0·6%)	45 (7·2%)	222 (8·9%)	10 (2·0%)	50 (2·1%)	21 (3·4%)	134 (3·6%)	28 (4·7%)	126 (4·5%)	11 (1·7%)	49 (1·7%)	5 (1·6%)	17 (1·7%)	128 (3·0%)	651 (3·2%)
Neonatal mortality	1 (0·2%)	4 (0·3%)	0 (0·0%)	0 (0·0%)	2 (0·3%)	4 (0·2%)	0 (0·0%)	0 (0·0%)	2 (0·3%)	9 (0·2%)	2 (0·3%)	4 (0·1%)	0 (0·0%)	0 (0·0%)	0 (0·0%)	1 (0·1%)	7 (0·2%)	22 (0·1%)
Male sex	206 (50·1%)	823 (51·6%)	319 (52·4%)	1861 (52·4%)	303 (48·5%)	1287 (51·6%)	250 (49·1%)	1173 (49·7%)	294 (47·6%)	1850 (50·0%)	305 (50·9%)	1471 (52·2%)	314 (49·1%)	1471 (50·1%)	158 (50·8%)	546 (53·2%)	2149 (49·7%)	10 482 (51·2%)
Exclusive breastfeeding at discharge	396 (96·4%)	1499 (94·0%)	418 (68·6%)	2870 (80·8%)	605 (96·8%)	2455 (98·5%)	387 (76·0%)	1720 (72·9%)	610 (98·9%)	3616 (97·7%)	577 (96·3%)	2736 (97·0%)	526 (82·2%)	2281 (77·6%)	267 (85·9%)	815 (79·4%)	3786 (87·6%)	17 992 (87·8%)
Mother admitted to intensive care unit	0 (0·0%)	3 (0·2%)	1 (0·2%)	2 (0·1%)	1 (0·2%)	1 (0·0%)	3 (0·6%)	7 (0·3%)	1 (0·2%)	5 (0·1%)	9 (1·5%)	18 (0·6%)	1 (0·2%)	1 (0·0%)	1 (0·3%)	1 (0·1%)	17 (0·4%)	38 (0·2%)
Birthweight ≥37^{+0} weeks (kg)	3·3 (0·4)	3·3 (0·4)	3·4 (0·4)	3·4 (0·4)	2·9 (0·4)	2·9 (0·4)	3·3 (0·4)	3·3 (0·4)	3·3 (0·4)	3·3 (0·4)	3·1 (0·4)	3·1 (0·4)	3·5 (0·5)	3·5 (0·5)	3·4 (0·4)	3·4 (0·5)	3·3 (0·4)	3·3 (0·5)
Birthlength ≥37^{+0} weeks (cm)	48·6 (1·7)	49·0 (1·7)	49·7 (1·6)	49·7 (1·6)	48·8 (1·8)	48·6 (1·8)	49·6 (1·8)	49·4 (1·7)	49·3 (1·8)	49·1 (1·8)	49·0 (1·8)	49·0 (1·8)	49·9 (1·9)	49·9 (1·9)	50·5 (2·0)	49·9 (2·2)	49·4 (1·9)	49·3 (1·8)
Birth head circumference ≥37^{+0} weeks (cm)	34·1 (1·1)	34·2 (1·2)	33·6 (1·2)	33·6 (1·2)	33·1 (1·2)	33·1 (1·1)	34·2 (1·2)	34·0 (1·2)	34·3 (1·2)	34·2 (1·2)	33·5 (1·1)	33·6 (1·1)	34·5 (1·3)	34·5 (1·3)	34·5 (1·4)	34·5 (1·4)	33·9 (1·3)	33·9 (1·3)

Data are n (SD) or n (%). Only pregnancies with a live singleton birth and no congenital malformation included. FGLS=Fetal Growth Longitudinal Study. NCSS=Newborn Cross-Sectional Study. PPROM=preterm prelabour rupture of membranes. NICU=neonatal intensive care unit.

Table 2: **Perinatal events for women enrolled in the FGLS and the FGLS-like subpopulation from the NCSS**

paternal size: the Indian parents were the shortest and those from the UK and USA the tallest, while mothers from Brazil, the UK, and the USA were the heaviest and Indian mothers the lightest (table 1). BMI was similar across the sites.

A detailed description of the environmental characteristics and working conditions in each area has been presented elsewhere.[16] Most women were not exposed to major environmental hazards that could affect pregnancy outcomes; housing conditions reflected the expected pattern for affluent populations in these communities in terms of size, water, sanitation, and construction characteristics.[16] Work outside the home varied substantially, indicative of the cultural patterns of middle to high socioeconomic status women in their respective countries. Of those (in this subsample of 555) who did work outside the house, 63% overall reported doing managerial, technical, or sales activities.

In the FGLS-like subpopulation from NCSS (table 1), the baseline characteristics across the eight sites were very similar to those of the FGLS population, which was as expected because women were selected from the underlying populations with the same criteria. Furthermore, women in the FGLS and FGLS-like populations were similar, except that fewer women in the FGLS-like subpopulation were nulliparous (63·4% *vs* 68·4% overall).

Baseline haemoglobin concentrations did not indicate that anaemia was a health issue in these populations at entry (table 1). Most women in FGLS and the FGLS-like subpopulation in low-income and middle-income countries were given iron and folic acid supplementation during pregnancy (range from 51% in Kenya to 100% in India; data not shown).

As expected, we recorded substantial variability in pregnancy and perinatal events across the sites (table 2). Preterm birth rates were low in both FGLS and the FGLS-like subpopulation of NCSS. There were similar sex ratios at birth in both populations without any major imbalances across the sites. Neonatal mortality up to hospital discharge was very low overall and per site in both populations, and most newborn babies were discharged from hospital on exclusive breastfeeding. These patterns provide confirmatory evidence of the adequate health and nutritional status of both study populations.

20 313 antenatal ultrasound scans were done in FGLS; the median number of scans (excluding the dating scan) was 5·0 (mean 4·9, SD 0·8) and 97% of women had four or more scans (mean 5·0, SD 0·6, range 4–7), which suggested that participants adhered well to the protocol (data not shown). 17 261 (85%) of the 20 313 ultrasound scans were done within the gestational age window requirements of the protocol (range from 76% in India to 93% in Oman).

For fetal head circumference measurements repeated during pregnancy, the within-site variance was seven times higher than the between-sites variance (table 3). The standardised site difference by gestational age for the eight sites was expressed as a proportion of the SD of all sites combined at each gestational age, for two ultrasound measurements: crown-rump length (9 weeks and 0 days to 13 weeks and 6 days of gestation) and fetal head circumference (≥14 and 0 days to 39 weeks and 6 days of gestation). The all sites' SD for crown-rump length ranged from 3·94 mm at 9 weeks and 6 days to 6·71 mm at 13 weeks and 6 days of gestation. For fetal head circumference, the all sites' SD ranged from 7·54 mm at 18 weeks and 6 days to 11·83 mm at 40 weeks and 0 days.

Within 10 fetal gestational age windows from 9 weeks and 0 days to 40 weeks and 6 days of gestation, representing 80 comparisons, 79 had values less than 0·5 (as prespecified in the protocol) of the SD of all sites combined. Crown-rump length ranged from –0·36 in India at 13 weeks and 0 days to 13 weeks and 6 days to 0·36 in Italy at 11 weeks and 0 days to 11 weeks and 6 days of gestation; fetal head circumference ranged from –0·58 in India (the only value outside the defined range) to 0·47 in Italy both between 34 weeks and 0 days and 40 weeks and 0 days of gestation (figure 2). Variability in crown-rump length remained constant over time but we noted a wider range of standardised site difference values for fetal head circumference with gestational age, mostly showing a smaller pooled SD for fetal head circumference rather than larger differences among countries (figure 2). On average, the head circumference at birth measurements across the eight sites were consistent with the fetal measurements (range –0·55 to 0·42; figure 2).

	Crown-rump length (N=4265)		Fetal head circumference (N=4237)		NCSS FGLS-like subpopulation birthlength† (N=20 166)	
	Estimate (SE)	Proportion	Estimate (SE)	Proportion	Estimate (SE)	Proportion
Variance between sites	0·65 (0·38)	1·9%	5·15 (2·82)	2·6%	0·12 (0·06)	3·5%
Variance between individuals within a site	‡	..	36·64 (1·54)	18·6%	‡	..
Residual variance	33·64 (0·73)	98·1%	155·70 (1·73)	78·8%	3·34 (0·02)	96·5%

FGLS=Fetal Growth Longitudinal Study. *Adjusted by gestational age as a fixed effect. †FGLS-like subpopulation represents the low-risk proportion of the total Newborn Cross-Sectional Study (NCSS) population selected with the same eligibility criteria as FGLS. Therefore, the subpopulation also included the newborn measures obtained from those enrolled in the FGLS cohort. ‡Variance between individuals within sites cannot be estimated because measures were collected only once per patient (cross-sectional).

Table 3: **Variance component analysis for crown-rump length, fetal head circumference, and newborn length***

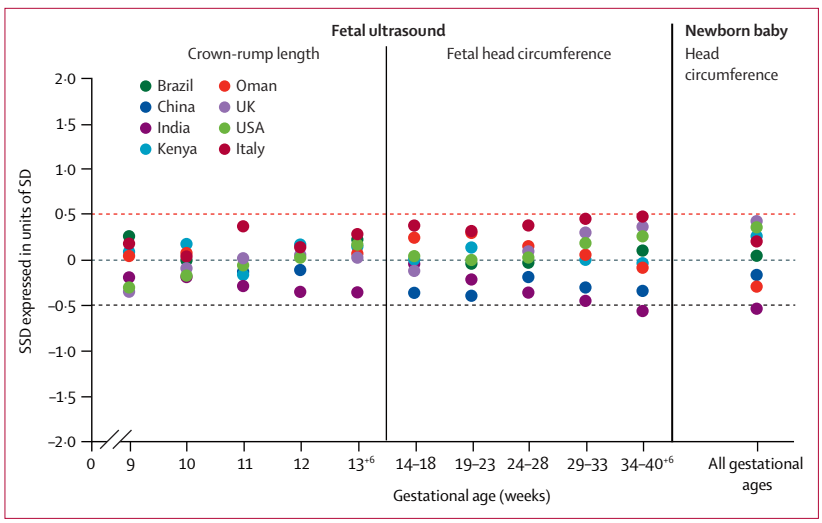

Figure 2: **Standardised site discrepancy for crown-rump length (N=4265), fetal head circumference (N=4237), and head circumference at birth (N=4217) in the Fetal Growth Longitudinal Study**
SSD calculated by: (site mean of either fetal head circumference, crown-rump length, or head circumference at birth minus all sites mean of either fetal head circumference, crown-rump length, or head circumference at birth at each gestational age interval)/all sites' SD of either fetal head circumference, crown-rump length, or head circumference at birth at each gestational age interval. SSD was adjusted at the median gestational age for all sites at each gestational age interval. SSD at birth were mean values by site across all gestational ages 25 weeks and 0 days to 43 weeks and 0 days calculated as above. The dashed red horizontal line is the 0·5 SD.[12] SSD=standardised site discrepancy.

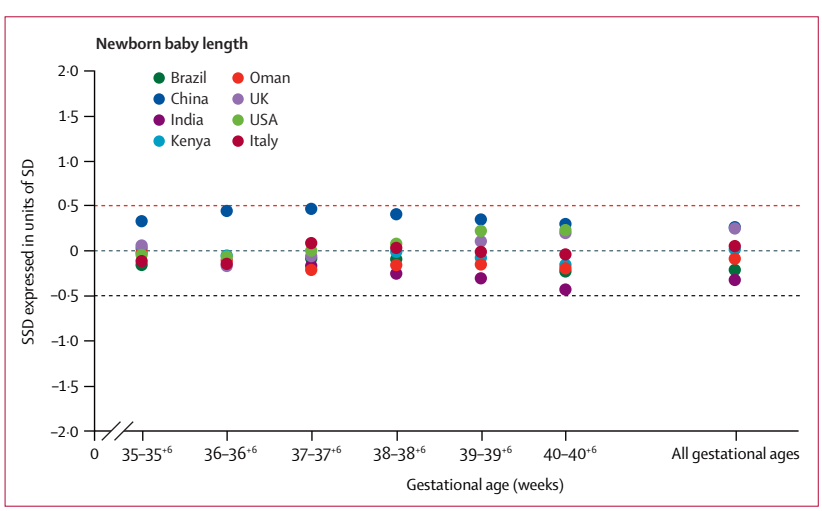

Figure 3: **Standardised site discrepancy of newborn length (N=20 166)**
The Fetal Growth Longitudinal Study (FGLS)-like subpopulation represents the low-risk proportion of the total Newborn Cross-Sectional Study (NCSS) population selected with the same eligibility criteria as FGLS, and it therefore includes all newborn babies in the FGLS cohort. SSD was the (site mean newborn length–all sites' mean newborn length at each gestational age interval)/all sites' SD of newborn length at each gestational age interval. Standardised site discrepancy was adjusted at the median gestational age for all sites at each gestational age interval. SSD at birth are mean values calculated by site across all gestational ages 25 weeks and 4 days to 44 weeks and 0 days. The dashed red horizontal line represents the 0·5 SD.[12] SSD=standardised site discrepancy.

To assess linear size at birth across sites, as in the WHO MGRS,[10] we did the same standardised site difference analysis for birthlength with all newborns in the FGLS-like subpopulation from NCSS, which included by definition all newborn babies from the FGLS cohort. Overall, the all sites' SD values ranged from 2·3 cm (at 35 weeks and 0 days to 35 weeks and 6 days of gestation) to 1·7 cm (at 40 weeks and 0 days to 40 weeks and 6 days of gestation). All standardised site difference values for the gestational age range for which we had an adequate sample size were within the prespecified –0·5 to 0·5 interval (figure 3). The standardised site difference values for birthlength across sites and gestational ages were all less than 0·5 of all sites' SD (range from –0·33 in India to 0·26 in China; figure 3).

In summary, we compared the eight sites with standardised site difference analysis during early pregnancy for crown-rump length (40 comparisons), late pregnancy for fetal head circumference (40 comparisons), and at birth for birthlength (48 comparisons): of these 128 comparisons, only one was marginally higher than 0·5 standardised site difference range. At birth, across gestational ages and sites, birthlength had standardised site difference values well below the lower than 0·05 threshold on average. The results of these two analyses show that the eight study populations are sufficiently similar in terms of skeletal size, on the basis of our predefined criteria, for the data to be pooled to estimate the 3rd, 50th, and 97th centiles for the pooled population.

From the pooled centiles (figure 4), we did sensitivity analyses for crown-rump length, fetal head circumference, and birthlength separately (the third step in our comparability analysis) to assess the effect on the centiles of the remaining pooled sample of removing a single site's data one at a time. We recorded no substantive effects on the remaining pooled sample's 3rd, 50th, and 97th centiles for any of the three primary measures in this exercise by individually removing each of the eight sites in our study from the pooled data. Additionally, we did the same analysis for head circumference at birth showing no substantive effect either.

As examples of these sensitivity analyses, figure 4 shows centile curves derived for each anthropometric measure, the effects of excluding, one at a time, the samples from China, India, Kenya, and the UK, whose general populations are usually believed to have very different sizes from one another. Removing data from these four different populations had no effect or only a minimal impact on the results from the remaining pooled sample. Exclusion of the other countries did not have an effect on the centiles either (data not shown). During data cleaning, we excluded 32 measures of head circumference at birth and 25 of birthlength because they were regarded as implausible within each study site's distribution or were more than 5 SD of the all sites' gestational age-specific mean.

Discussion

We have presented data obtained under rigorously controlled methods, comparing fetal skeletal growth and newborn baby and infant sizes from 9 weeks of gestation to birth, in healthy, well nourished women living in environments with minimal constraints on fetal growth,

across eight geographically diverse urban areas worldwide (panel). We selected fat-free mass (ie, skeletal) indicators as the primary measurements to compare fetal growth and newborn size across the study sites. These measurements are recommended as the best option to compare growth among populations[10] because they are resistant to skewing in response to excessive nutrition, as opposed to weight or other fat-related indicators; although they could be affected by undernutrition or infection, these factors are unlikely to play an important part in our healthy populations; they are normally distributed (unlike fat-related indicators); are more precise than fat-related measures especially for fetal ultrasound; and were used to compare populations in the aforementioned WHO MGRS and to construct the infant and child growth standards currently in use, thereby providing conceptual continuity with our project.[10,12,23]

With variance component analysis, we showed that only between 1·9% and 3·5% of the total variability in fetal skeletal growth and newborn length could be attributed to between-site differences. This is remarkably similar to the 3% variability reported by both the WHO MGRS for infant length,[10] and Habicht and colleagues[7] for child height.

We used a prescriptive approach to select the study populations to assess the growth potential of the fetuses and newborns—ie, we studied women at low risk of fetal growth disturbances because of both their individual characteristics selected according to a predefined set of criteria[15] and the socioeconomic and demographic characteristics of the underlying populations, in settings with diverse ethnic backgrounds as recommended by the WHO Expert Committee on Physical Status: the use and interpretation of anthropometry in 1995.[31] This strategy allows fair comparisons across populations where the health and nutrition needs of mothers are met and adequate, standardised antenatal care is provided. For the purpose of studying optimum fetal growth, having a representative sample of whole cities is not desirable, especially in low-income countries, and would conflict with a prescriptive approach.[9] The term prescriptive approach is used in the scientific literature to describe the process of production of biological norms to be

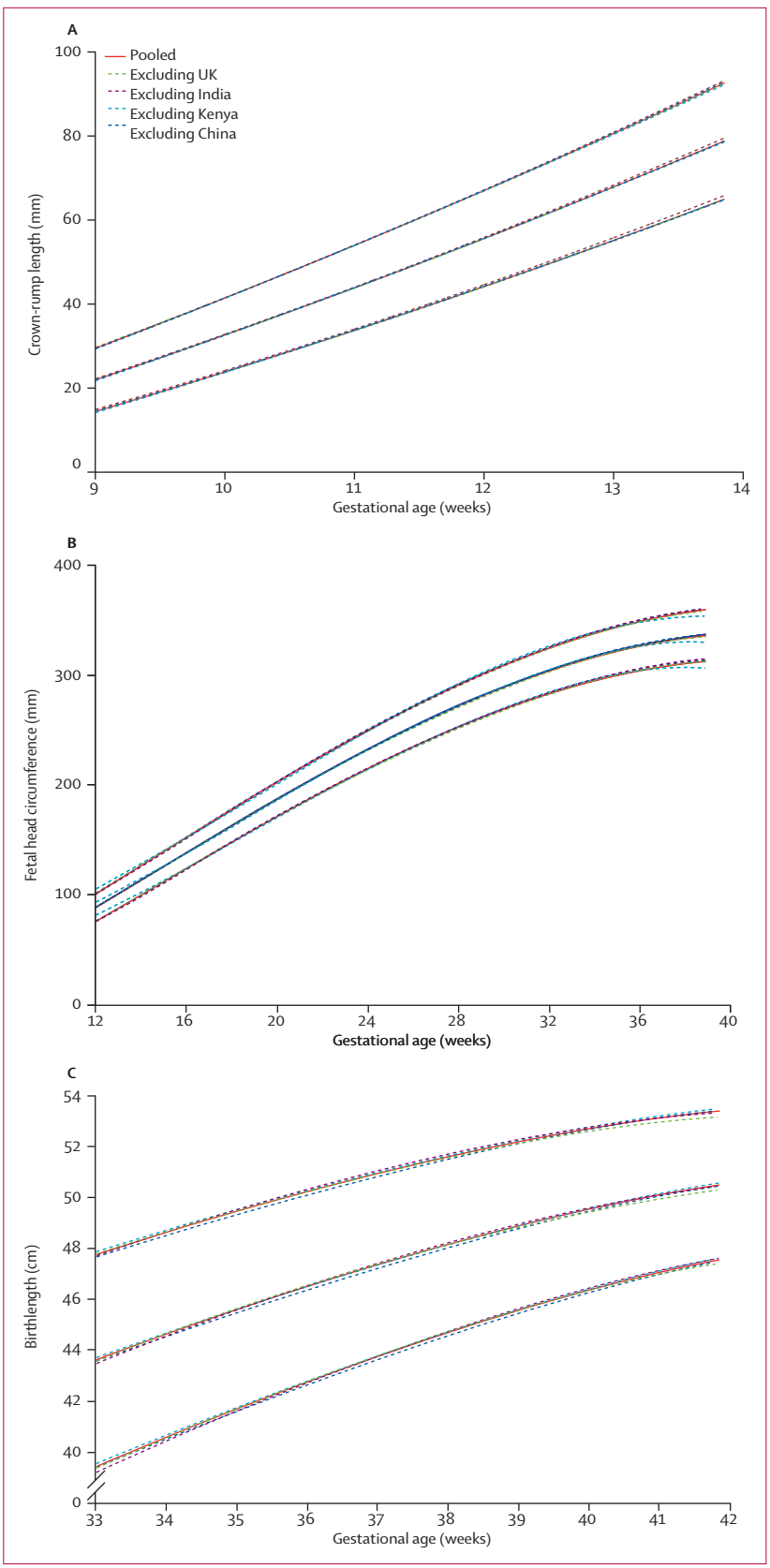

Figure 4: **Sensitivity analyses**

(A) Crown-rump length at 3rd, 50th, and 97th centiles estimated with fractional polynomial regression models[30,33] for the total Fetal Growth Longitudinal Study (FGLS) population (N=4265, solid line) and the remaining sample after data from the sites in China, India, Kenya, and the UK, were excluded. (B) Fetal head circumference at 3rd, 50th, and 97th centiles estimated with fractional polynomial regression models[30,33] for the total FGLS population (N=4237, solid line) and the remaining sample after data from the sites in China, India, Kenya, and the UK were excluded. (C) Birthlength at 3rd, 50th, and 97th centiles estimated with fractional polynomial regression models[30,32] for the FGLS-like subpopulation, which represents the low-risk proportion of the total Newborn Cross-Sectional Study (NCSS) population (N=20 166) selected with the same eligibility criteria as FGLS (solid line), and the remaining sample after data from the sites in China, India, Kenya, and the UK were excluded. The FGLS-like subpopulation included the FGLS cohort.

Panel: Research in context

Systematic review

Optimum intrauterine growth in populations with minimum identifiable risk factors for adverse maternal and perinatal outcomes has not previously been defined. International fetal skeletal growth and newborn linear size standards, developed with a prescriptive approach as in the WHO Multicentre Growth Reference Study (MGRS) for infants and children, are not available. Recent systematic reviews[13,14] strongly support the need for international growth standards for the prenatal and neonatal periods, and their development was the main objective of the INTERGROWTH-21st Project. To produce such standards, similarities of growth across geographically diverse populations need to be shown.

Interpretation

Fetal skeletal growth and newborn linear size are strikingly similar among geographically diverse populations when mothers' environmental, health, and nutritional conditions are met. Our data suggest that differences in these measures reported in the scientific literature are more likely due to environmental and socioeconomic differences than genetic variation, as has been shown for infants and children. The pooled data from the eight study sites are being used to construct international standards for fetal skeletal growth and newborn linear size, which will complement the existing standards generated by the WHO MGRS for infants and children, and allow growth to be monitored at individual and population levels from early pregnancy to 5 years of age.

achieved or aspired to at individual and population levels (to construct growth standards) as opposed to a reference, which only describes the distribution of fetal or newborn anthropometric variables in one population at a given time and place.

The project had a few limitations related to implementation of such research across maternity units, which did not all adhere to uniform clinical guidelines. Recruitment and study of almost 60 000 pregnant women at eight very different sites, including standardisation of the clinical practice of more than 300 health professionals, and monitoring of the quality of their data, was a major challenge. This limitation is reflected in some of the differences among countries: for example, in the FGLS-like subpopulation derived from NCSS, the mean gestational age at the first ultrasound scan ranged from 12·0 (SD 4·0) weeks of gestation in the USA to 17·1 (7·9) weeks of gestation in Kenya and the preterm birth rate ranged from 10·0% in India to 3·4% in the UK, which resulted in an unbalanced sample size at the lowest gestational ages (<32 weeks and 0 days of gestation) in the newborn baby comparisons. Nevertheless, these differences across sites had a positive effect by increasing variability in the data (an important feature for a screening method), ensuring that the growth standards being produced[32] will be suited to the diverse characteristics of the populations in which they will be used.

Furthermore, in some of the study populations (eg, India and Oman), participating women were shorter than those in the other countries despite their affluence and high educational attainments; fetal head circumference and head circumference at birth from these populations were in the lower range of the overall study population, yet within the pre-established bounds of similarity. The findings of the sensitivity analyses are reassuring because exclusion of these samples did not change the pooled extreme centiles, which is of huge practical importance for the global implementation of the standards to be generated from these data.

Conversely, and potentially of greater biological significance, overall birthlength (the main skeletal growth measure to compare newborn babies across populations, as in the WHO MGRS) was very similar among all sites, and the mean values are closer to the other sites than the corresponding maternal height (figure 3). This finding is very relevant because it has been recently reported, with data from the WHO MGRS cohort at 2 years of age, that the predicted adult height for the same sites was 7 cm greater than the mid-parental height, suggesting large improvements in stature in just one generation.[33]

Nevertheless, as we recognised in the protocol,[15] some variability in these populations remained, mostly at the extremes of gestational age in some parameters. This variation might have arisen because of residual secular trends, true inter-ethnic differences,[10] unstable estimations due to the small sample sizes at some gestational age windows, or simply differences in protocol implementation despite our best efforts to standardise rigorously across the study sites. However, we confirmed that such variability among sites represents only 3% of the total variance for skeletal growth, whereas the variability in individuals within a site is seven times higher (table 3). This finding addresses our a priori question: is the variability across populations for the three primary size measures larger than the variability of the same measures within populations?[15] Lastly, the study patients did not undergo genetic profiling and, although this might seem to be a limitation, the eight populations included in the study are unlikely to be homogeneous when compared with each other.

The INTERGROWTH-21st Project has features that are unique in the specialty of fetal and infant growth research: first, it was conceptually and methodologically very similar to the WHO MGRS, even sharing two study sites (Pelotas, Brazil; Muscat, Oman) and two other countries (India and the USA); second, it was population-based according to geographical and socioeconomic criteria,[34] including more than 80% of all the deliveries in these well defined populations; third, it included geographically diverse populations across the world; fourth, we standardised the measurement of fetal size with centrally trained staff and the same ultrasound machine that was specially adapted

to allow masking; fifth, we developed a novel, quality control strategy for ultrasound measures; and sixth, similar care was taken with the newborn baby and infant anthropometric measures by use of the same protocols, equipment, and quality control methods across sites as in the WHO MGRS. Furthermore, the FGLS population was very similar to the underlying FGLS-like subpopulation (roughly 34% of the NCSS cohort). This finding suggests that no major bias existed in selection of the FGLS cohort from the total number of eligible women, and provides strong external validity of the results.

Importantly, the study populations presented here are consistent at birth with those in the WHO MGRS. For example, the mean birthlength at greater than 37 weeks and 0 days of gestation in the WHO MGRS was 49·5 cm (SD 1·9),[10] which was very similar to that in the FGLS cohort and the NCSS FGLS-like subpopulation (table 2). Furthermore, the mean birthweights in FGLS and the NCSS FGLS-like subpopulations were identical to the WHO MGRS birthweight of 3·3 kg (0·5).[10] What is even more remarkable in biological terms perhaps is that, at 1 year of age, the FGLS cohort is on the 49th (boys) and 52nd (girls) centiles for length and the 49th (boys) and 50th (girls) centiles for head circumference on the WHO Child Growth Standards (data not shown).[12]

These clinical and anthropometric similarities in the study populations are compatible with recent genetic findings showing that, when distinct populations are considered (eg, sub-Saharan Africans and Europeans) and hundreds of genetic loci are analysed, individuals are frequently more similar to members of other populations than to those in their own population.[35] These comparisons show similarities in skeletal size up to 1 year of age in two cohorts of geographically and ethnically diverse populations, but selected with the same environmental, health, and nutritional criteria. They provide reassurance that the cohort of fetuses and newborns we have studied in the INTERGROWTH-21st Project was healthy and well nourished according to the WHO MGRS criteria. These results support pooling of the data for the construction of international standards. The data are in strong agreement with those of the WHO MGRS, and suggest that differences reported in the scientific literature in fetal growth and newborn size are more likely due to environmental and socioeconomic differences than genetic variation, as has been shown for infants and children.

Contributors

JV and SHK were responsible for the idea and conception of the project. JV, SHK, DGA, and JAN prepared the original protocol, with later input from ATP, LCI, FCB, and ZAB. JV, ATP, LCI, AL, and ZAB supervised and coordinated the project's overall undertaking. EOO, DGA, FCB, and CCV were responsible for data management and analysis in collaboration with JV. RP, FCB, MC, YAJ, EB, MGG, MP, and IOF collaborated in the project and implementation in their respective countries. JV and SHK wrote the report with input from all coauthors. All coauthors read the report and made suggestions on its content.

Declaration of interests

We declare no competing interests.

Acknowledgements

This study was supported by a grant (49038) from the Bill & Melinda Gates Foundation to the University of Oxford. We thank the Health Authorities in Pelotas, Brazil; Beijing, China; Nagpur, India; Turin, Italy; Nairobi, Kenya; Muscat, Oman; Oxford, UK; and Seattle, USA, who helped with the project by allowing participation of these study sites as collaborating centres. We also thank Philips Healthcare for providing the ultrasound equipment and technical assistance throughout the project; MedSciNet UK for setting up the INTERGROWTH-21st website and for the development, maintenance, and support of the online data management system; the parents and infants who took part in the studies and the more than 200 members of the research teams who made the implementation of this project possible. The participating hospitals included: Brazil, Pelotas (Hospital Miguel Piltcher, Hospital São Francisco de Paula, Santa Casa de Misericórdia de Pelotas, and Hospital Escola da Universidade Federal de Pelotas); China, Beijing (Beijing Obstetrics and Gynecology Hospital, Shunyi Maternal & Child Health Centre, and Shunyi General Hospital); India, Nagpur (Ketkar Hospital, Avanti Institute of Cardiology Private Limited, Avantika Hospital, Gurukrupa Maternity Hospital, Mulik Hospital and Research Centre, Nandlok Hospital, Om Women's Hospital, Renuka Hospital & Maternity Home, Saboo Hospital, Brajmonhan Taori Memorial Hospital, and Somani Nursing Home); Kenya, Nairobi (Aga Khan University Hospital, MP Shah Hospital, and Avenue Hospital); Italy, Turin (Ospedale Infantile Regina Margherita Sant' Anna and Azienda Ospedaliera Ordine Mauriziano); Oman, Muscat (Khoula Hospital, Royal Hospital, Wattayah Obstetrics & Gynaecology Poly Clinic, Wattayah Health Centre, Ruwi Health Centre, Al-Ghoubra Health Centre, and Al-Khuwair Health Centre); UK, Oxford (John Radcliffe Hospital), and USA, Seattle (University of Washington Hospital, Swedish Hospital, and Providence Everett Hospital). Full acknowledgment of all those who contributed to the development of the INTERGROWTH-21st Project appears online.

References

1 Victora CG, Adair L, Fall C, et al. Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* 2008; **371:** 340–57.
2 Goedhart G, van Eijsden M, van der Wal MF, Bonsel GJ. Ethnic differences in term birthweight: the role of constitutional and environmental factors. *Paediatr Perinat Epidemiol* 2008; **22:** 360–68.
3 Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1987; **65:** 663–737.
4 Frayling TM, Hattersley AT. The role of genetic susceptibility in the association of low birth weight with type 2 diabetes. *Br Med Bull* 2001; **60:** 89–101.
5 Bhandari N, Bahl R, Taneja S, de Onis M, Bhan MK. Growth performance of affluent Indian children is similar to that in developed countries. *Bull World Health Organ* 2002; **80:** 189–95.
6 Owusu WB, Lartey A, de Onis M, Onyango AW, Frongillo EA. Factors associated with unconstrained growth among affluent Ghanaian children. *Acta Paediatr* 2004; **93:** 1115–19.
7 Habicht JP, Martorell R, Yarbrough C, Malina RM, Klein RE. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? *Lancet* 1974; **1:** 611–14.
8 WHO Multicentre Growth Reference Study Group. Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. *Acta Paediatr Suppl* 2006; **450:** 7–15.
9 Garza C, de Onis M. Rationale for developing a new international growth reference. *Food Nutr Bull* 2004; **25** (suppl 1): S5–14.
10 WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. *Acta Paediatr Suppl* 2006; **450:** 56–65.
11 de Onis M. Update on the implementation of the WHO child growth standards. *World Rev Nutr Diets* 2013; **106:** 75–82.
12 de Onis M, Garza C, Onyango AW, Martorell R. WHO child growth standards. *Acta Paediatr* 2006; **450:** 1–101.
13 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.
14 Napolitano R, Dhami J, Ohuma E, et al. Pregnancy dating by fetal crown-rump length: systematic review of charts. *BJOG* 2014; **121:** 556–65.

15 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.

16 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.

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

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

19 Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ* 2012; **345:** e7976.

20 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.

21 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.

22 Cheikh Ismail L, Knight HE, Bhutta Z, et al. Anthropometric protocols for the construction of new international fetal and newborn growth standards: the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 42–47.

23 de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. *Food Nutr Bull* 2004; **25** (suppl 1)**:** 27–36.

24 Cheikh Ismail L, Knight H, Ohuma E, et al. Anthropometric standardisation and quality control protocols for the construction of new, international, fetal and newborn growth standards: the INTERGROWTH-21st Project. *BJOG* 2013; **120** (suppl 2)**:** 48–55.

25 Bhutta Z, Giuliani F, Haroon A, et al. Standardisation of neonatal clinical practice. *BJOG* 2013; **120** (suppl 2)**:** 56–63.

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

27 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.

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

29 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.

30 Silverwood RJ, Cole TJ. Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound Obstet Gynecol* 2007; **29:** 6–13.

31 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.

32 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* (in press).

33 Garza C, Borghi E, Onyango AW, de Onis M, WHO Multicentre Growth Reference Study Group. Parental height and child growth from birth to 2 years in the WHO Multicentre Growth Reference Study. *Matern Child Nutr* 2013; **9** (suppl 2)**:** 58–68.

34 Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998; **20:** 81–90.

35 Witherspoon DJ, Wooding S, Rogers AR, et al. Genetic similarities within and between human populations. *Genetics* 2007; **176:** 351–59.