

# The INTERGROWTH-21st Fetal Growth Standards: toward the global integration of pregnancy and pediatric care

Aris T PAPAGEORGHIU MD<sup>1</sup>, Stephen H KENNEDY MD<sup>1</sup>, Laurent J SALOMON MD<sup>2</sup>, José VILLAR MD<sup>1</sup> for the International Fetal and Newborn Growth Consortium for the 21<sup>st</sup> Century (INTERGROWTH-21<sup>st</sup>)

<sup>1</sup> Nuffield Department of Obstetrics & Gynaecology and Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK;

<sup>2</sup> Department of Obstetrics and Fetal Medicine, Hopital Necker Enfants Malades, Université Paris Descartes, Paris, France.

**STATEMENT OF FINANCIAL SUPPORT:** This study was supported by the INTERGROWTH-21<sup>st</sup> grant 49038 from the Bill & Melinda Gates Foundation to the University of Oxford. The funder played no role in the study design: collection, analysis and interpretation of data; the writing of the report or the decision to submit the article for publication.

**Conflict of interest Disclosures:** The authors report no conflicts of interest.

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

## Abstract

The purpose of the INTERGROWTH-21<sup>st</sup> Project was to develop international, prescriptive standards for fetal growth assessed by ultrasound and fundal height, preterm postnatal growth, newborn size and body composition, maternal weight gain and infant development at the age of 2. Hence, we have produced, based on WHO recommendations, the first comprehensive set of international standards of optimal fetal and newborn growth that perfectly match the existing WHO Child Growth Standards. Uniquely, the *same population* was followed longitudinally from 9 weeks of fetal life to 2 years of age, with growth, health and nutritional status assessment at 2 years supporting the appropriateness of the population for construction of growth standards. The resulting package of clinical tools allows, for the first time, growth and development to be monitored from early pregnancy to infancy. The INTERGROWTH-21<sup>st</sup> Fetal Growth Standards, which are based on observing over 4500 healthy pregnancies, nested in a study of over 59000 pregnancies from healthy populations, show how fetuses should grow - rather than the more limited objective of past references, which describe how they have grown at specific times and locations. Our work has confirmed the fundamental biological principle that human growth in populations is mostly dependent on environmental, nutritional and socio-economic factors. Thus, when mothers' nutritional and health needs are met, and environmental constraints on growth are low, we found that less than 3.5% of the total variability of skeletal growth was due to differences between sites. ~~Not~~ We propose that not recognizing the concept of optimal

growth could deprive the most vulnerable mothers and their babies of optimal care, because local growth charts normalize those at highest risk for growth restriction and overweight, and limits the amount of information policymakers need for rigorous evaluation and effective resource allocation. We strongly encourage colleagues to join efforts to provide integrated, evidence-based growth monitoring to pregnant women and their infants worldwide. Presently, there are 23.3 million infants born small for gestational age in low- to middle-income countries according to the INTERGROWTH-21<sup>st</sup> Newborn Size Standards. The misclassification of these infants by using local charts could affect the delivery of optimal healthcare to them.

Key words: Fetal Growth Standards; INTERGROWTH-21<sup>st</sup>; pregnancy

## Introduction

Recent publications<sup>1-5</sup> and ensuing editorials and correspondence,<sup>6-9</sup> as well as presentations and debates at national and international meetings, have activated a controversy that goes well beyond the boundaries of obstetrics and perinatal medicine. The controversies touch upon fundamental topics in biology, genetics, politics and human rights. Sadly, some arguments are reminiscent of the historical dispute about the influence of race/ethnicity/skin color on human head size and shape.<sup>10</sup>

There is little disagreement about the uniformity of human growth across healthy populations in early pregnancy, and the applicability of international standards to estimate gestational age,<sup>2, 11</sup> evaluate size at birth worldwide,<sup>12-14</sup> and monitor the growth of term newborns up to 5 years of age.<sup>15</sup> Unfortunately, however, challenges are being made to key conceptual and factual issues relating to fetal growth monitoring in the second half of pregnancy that are preventing the introduction of integrated care across the first 1000 days of life.

Some members of the obstetric community seem to hold firmly to the view that fetal growth differences among healthy populations, specifically after 14 weeks' gestation, are strongly influenced by maternal factors such as self-reported ethnicity/nationality or political borders. This position is difficult to sustain given the now strong evidence, obtained from detailed monitoring of low-risk cohorts from early pregnancy to 2 years of age, that human growth, evaluated by markers of skeletal, fat-free mass (i.e. fetal head circumference (HC), birth length, HC at birth and infant length) is very similar amongst low-risk populations regardless of where they live, or their race/ethnicity/skin color,<sup>16, 17</sup> as demonstrated more than a decade ago by the WHO Multicentre Growth Reference Study (MGRS).<sup>18</sup>

Differences observed in perinatal health among general populations across countries are principally due to the downstream effects of environmental, nutritional and socio-economic factors and this has important consequences. These are well-recognized in medicine and public health, i.e. a mother's zip code is a better indicator of her health status than her genetic code.<sup>19, 20</sup> Our aim here, therefore, is to dispel these misconceptions and

unsubstantiated beliefs that, if left uncorrected, could adversely affect the quality of care offered to women and their families.

## Methodological issues relevant for the screening of fetal growth abnormalities in the general pregnant population

### References versus standards

At present, clinicians are using a large number of different ultrasound charts of fetal size, based on a variety of populations and methodologies, to monitor growth. However, in a series of systematic reviews, we have shown that the overwhelming majority of these charts were developed with important methodological flaws.<sup>21-23</sup>

All these charts are references rather than ‘prescriptive’ standards. The distinction is critical. References describe how individuals have grown at a particular time and place usually decades beforehand. Prescriptive standards, on the other hand, which are purposely developed using a selected, healthy population, describe how humans *should* grow when nutritional, environmental and health constraints on growth are minimal. They are based conceptually on the WHO 1995 recommendation that “human growth should be evaluated using international standards, describing how individuals should grow”.<sup>24</sup> Results from any screening test, so also in the case of growth monitoring using a standard, requires clinical judgement to interpret findings and determine future actions.

The use of references instead of standards has important implications at individual and populations levels that impact directly on clinical care and public health policies. To understand why, it is important to realize that the distribution of size in the general population does not constitute a standard. The prevalence of stunting amongst children globally illustrates the point well as the rate is inversely related to the level of socioeconomic status (SES).<sup>25</sup> Therefore, growth charts based on the distribution of biometric measures in low and high SES populations will be very different from each other. A chart based on a low SES sample will clearly underestimate the prevalence of small-for-gestational age (SGA) and stunting, which are markers of social inequity.<sup>25</sup>

These differences can be illustrated when assessing the INTERGROWTH-21<sup>st</sup> Project and the WHO sponsored study by Kiserud et al,<sup>4</sup> which had completely different objectives. The former was a comprehensive evaluation of human growth and development across the first 1000 days of life, leading to the construction of fetal and preterm postnatal growth standards; it included an assessment of newborn body composition, infant feeding practices, and preterm postnatal growth, as well as postnatal growth and neurodevelopment evaluation at 2 years of age to assess the appropriateness of the complete cohort for the construction of standards (Panel 1). The INTERGROWTH-21<sup>st</sup> Project<sup>26</sup> also adhered rigorously to the WHO recommendations for assessing human size and growth (see below).<sup>24</sup> In contrast, the paper by Kiserud et al. presented a hospital-based study, which generated fetal growth references not standards<sup>4</sup>; the selection of the population to study, outcome measures, ultrasound equipment and analytical strategy were different, as indeed was the lack of masking the ultrasound measures and thereby avoid potential observer bias.

The need to differentiate standards from reference charts is not an ethereal intellectual matter but a vitally important global issue with marked political and socioeconomic ramifications. How else can progress towards UN Sustainable Developmental Goal 3.1 (end preventable deaths of newborns and children under 5 years of age) be measured unless international standards are used to compare the health and nutritional status of infants, as was done in assessing progress towards Millennium Development Goal 1 (eradicate extreme poverty and hunger) by showing changes in stunting rates based on the international WHO Child Growth Standards.<sup>27</sup> Making late fetal growth charts country- or region-specific would not only make this task impossible but would confuse the interpretation of all other growth and health indicators across populations.

### How were the INTERGROWTH-21<sup>st</sup> populations selected?

The first step in creating prescriptive international standards of optimal fetal growth was to select free-living populations in defined geographic areas with minimal constraints on growth, and good maternal and perinatal health outcomes. The second step was to select, from the whole population, healthy pregnant women at low risk of adverse outcomes.<sup>26</sup> This is completely different to the policy of recruiting women from selected hospitals, which frequently introduces bias because women who attend certain hospitals for pregnancy care may be different from the overall population of pregnant women - particularly when the population is served by private and public hospitals, and recruitment is from one but not the other.

In the INTERGROWTH-21<sup>st</sup> Project (Panel 2), all institutions providing obstetric care in eight delimited urban areas with no or low levels of major, known, non-microbiological contamination,<sup>28</sup> were selected to ensure that the study was population-based. Thus, a strategy of including delimited geographic areas where the health, educational and nutritional needs of all the inhabitants are mostly reached is very different to that of the WHO sponsored<sup>4</sup> and NICHD<sup>5</sup> fetal studies, which simply selected a number of convenient hospitals. Indeed, the lack of population-based sampling means that the sites selected by these two studies would have been ineligible for the INTERGROWTH-21<sup>st</sup> Project.

Healthy pregnant women with a naturally conceived singleton pregnancy, who met the individual inclusion criteria,<sup>26</sup> were identified prospectively in the INTERGROWTH-21<sup>st</sup> Project. Approximately one third of the healthy women who met these criteria were enrolled in the Fetal Growth Longitudinal Study (FGLS), one of the project's five studies. This is the cohort from which-the Fetal Growth Standards,<sup>1</sup> and the standards for: i) symphysis-fundal height,<sup>29</sup> ii) gestational weight gain,<sup>30</sup> iii) early pregnancy dating,<sup>2</sup> iv) estimated fetal weight (EFW),<sup>3</sup> v) newborn body composition,<sup>31</sup> and vi) the postnatal growth of preterm infants were derived,<sup>32</sup> as well as, in 2018, vi) fetal velocity growth and vii) neurodevelopment at 2 years of age. Underlying these tools was a series of systematic reviews of current clinical practice and development of methodologies based on a deep understanding of the issues in order to arrive at optimal scientific analytical framework. It should be noted that there was wide heterogeneity in methods, tests and definitions used. (Panel 3).

Women were recruited before 14 weeks' gestation, and pregnancies were dated based on a certain last menstrual period, but corroborated by ultrasound measurement of the crown-rump length (CRL).<sup>33</sup> Ultrasound scans were then performed every 5±1 weeks from the initial dating scan by dedicated research staff using identical, mid-range ultrasound machines at each study site, with rigorous training and standardization procedures,<sup>34, 35</sup> quality control measures,<sup>36</sup> and blinding of measurements.

Moreover, unlike any other longitudinal study of ultrasound in pregnancy, the infants involved in the Fetal Growth Standards were followed for 2 years after birth, using the same standardized methods employed in the WHO Child Growth Standards to measure growth,<sup>18</sup> neurodevelopment, auditory processing and sleep-wake patterns at 2 years of age.<sup>37</sup> We have recently reported that the Fetal Growth Standards cohort remained healthy up to 2 years of age, with adequate growth and motor development assessed using WHO tools,<sup>15, 38</sup> supporting its appropriateness for the construction of the international fetal and preterm postnatal growth standards.<sup>17</sup>

### Evaluating similarities in fetal growth among populations

Critics of the INTERGROWTH-21<sup>st</sup> Project often misquote our conclusions by claiming we believe that “all babies everywhere grow in the same way” or “birth weight is the same in general populations throughout the world”. This is self-evidently not the case. Rather we have proved that **measures of fetal and newborn skeletal growth are similar across diverse geographical settings when most of the mothers' socioeconomic, educational, nutritional and health needs are met and environmental constraints on growth are low.**<sup>16</sup> This conceptual approach is far more comprehensive than just selecting women without clinical risk at hospital level, and is the key justification for the construction and use of international fetal growth standards.

Skeletal growth was chosen as the outcome measure to evaluate similarities in growth based on the WHO recommendation to avoid fat-based indicators, e.g. abdominal circumference (AC), when comparing populations for the construction of human growth standards. This is vitally important for fetal growth screening in developed countries with an obesity epidemic, and for those developing countries in “epidemiological transition”. The specific recommendation is to use markers of skeletal or linear growth because they are: a) mostly resistant to skewing in response to “excessive nutrition”<sup>39</sup>; b) normally distributed (unlike fat-related indicators); c) more precisely measurable than fat-related indicators; d) consistent with pediatric practice worldwide as they were used by WHO to generate the WHO Child Growth Standards, and e) although responsive to under-nutrition or infection, this is hardly relevant in our healthy population. The comparison to assess similarities or differences in the MGRS was based on infant height. The corresponding measure in fetuses is crown-rump length that can be measured reliably until 14 weeks' gestation and again at birth; these showed no differences amongst sites using the three analytical approaches described below. It is difficult to see that large variation should exist between these two time-points, but we tried to assess it; as length is not measurable (due to fetal curling) we used HC as a skeletal measure.

Conversely, the use of Estimated Fetal Weight (EFW), a composite calculation of three different fetal anthropometric parameters, contradicts these physiological concepts and adds considerable error to any estimation, especially at term. In addition, fetuses can reach the same EFW through several permutations of the equation's components, which are clearly not comparable. This old equation,<sup>40</sup> determined using a very small sample (although it is difficult to identify the exact sample size from the published papers) adds to the confusion, especially as it includes femur length (FL) measures obtained using old ultrasound equipment, which yields different results to modern ultrasound machines.<sup>41</sup>

Many studies have shown that the 95% confidence interval of the random error associated with EFW accuracy exceeds 14% of birth weight, which is close to 400g for the average birth weight at term. In fact, a systematic review concluded that "the size of the random errors (of EFW) remains a major obstacle to confident use in clinical practice".<sup>42</sup> Furthermore, accuracy is even worse for small and large fetuses for whom growth estimation is clinically more important. These are very important issues when comparing differences in EFW values amongst populations, which is why it is much more logical to compare populations using the individual skeletal parameters, such as HC and FL, separately.

Then, why not to use AC alone as a parameter to compare populations considering that it is associated with perinatal outcomes in late pregnancy? The response, based on the recommendation of skeletal linear growth, is that AC is a fat/tissue-based measure equivalent to weight. Hence, if a marker of fat/tissue mass were used to compare growth across populations, it would be observed in the third trimester of pregnancy that, compared to non-overweight populations, those in the midst of the obesity epidemic have fetuses with abnormal larger AC values.-This is observed in one of the NICHD publications,<sup>43</sup> which shows an increase in AC in overweight women compared to those of normal weight.

Importantly, in the context of the NICHD argument, fetal AC changes are mostly due to liver growth supported by a small component of abdominal subcutaneous fat. It is, therefore, very difficult to understand how the different ethnic groups they studied, can have differential, genetically driven, liver growth during the second part of pregnancy.

### **Why outcome measures should be masked**

It is a basic research principle across all biological subjects that any outcome measures being obtained, especially by medical observers, should be masked as prior knowledge or real time plotting by the operator can strongly influence their measurement. This is averred of the hypothesis under investigation, which certainly applies to ultrasound parameters as the operator can strongly influence their measurement. In the INTERGROWTH-21<sup>st</sup> Project, the identical ultrasound machines used at every site were adapted to enable measurements to be taken in a blinded fashion; this has not been the case in all other fetal studies despite the well-recognized potential for bias.

### **Which is the most appropriate analytical strategy?**

Data from studies comparing populations for creating international standards should be analyzed in two steps: first, by evaluating the similarities among sites (or 'ethnic' groups in

the case of the NICHD fetal study) and second by estimating the centiles. For the first step, we followed the WHO recommended strategy for the construction of growth standards based on three complementary methods after the literature was systematically reviewed;<sup>44</sup> the WHO sponsored study simply used a p value-multiple testing based strategy; while the NICHD study was designed to create standards for the four ethnic groups, so the issue of potential pooling was not assessed. In the NICHD study, judgements of differences were again based on a p-value and although the potential clinical significance was assessed, this was done a posteriori rather than as a judgement on whether pooling should take place or not. This is another core element that differentiates these three publications (Panel 1).

It is also obvious that statistical significance is not the same as clinical significance and that small differences, well within the measurement error of ultrasound equipment, may become statistically significant in a study with a large sample size. For example, in a study based on the women enrolled in the NICHD study,<sup>43</sup> the median femur length was 0.8mm longer and the median humerus length 0.6mm longer in 443 obese versus 2320 nonobese women (in context the lengths were 71.0 vs 70.2 mm; and 62.2 vs 61.6 mm, respectively). The differences are judged to be statistically significant because the p values are 0.01 and 0.03, respectively, but are of minimal clinical relevance and well within measurement error, which may in any case be higher in obese women.

Therefore, the clinically relevant difference should be defined *a priori*, and p-values have no place in deciding whether these differences are clinically significant. Furthermore, the use of multiple testing for comparing individual populations against each other is an irrelevance, as it was never suggested that fetuses from one site, e.g. in India, should be evaluated using charts from another site, e.g. in Brazil or USA. What is recommended is to create international standards by combining data from prescriptive populations against which all samples are compared. Standardization of tools and measures is a central practice not only in biology and medicine, but also in all fields of science and even the arts.

Hence, the INTERGROWTH-21<sup>st</sup> Project set up its *a priori* conditions on whether to pool the datasets based on the WHO internationally accepted three-component strategy used for the WHO Child Growth Standards. The clinically relevant difference was also defined *a priori* using the appropriate, recommended outcome measure for evaluating growth across populations. Therefore, it seems odd that investigators, who must have been aware of the WHO expert committee recommendations, did not follow them or use even the most practical analytical method, namely sensitivity analysis. This is highly relevant because they advocated pooling their data, despite calling them different, without conducting a standard sensitivity analysis.<sup>4</sup>

To conduct such an evaluation, we used the SD of the all sites' combined value of fat-free measures as the denominator for the standardization process,<sup>45</sup> following WHO previous work.<sup>18</sup> For fetal growth, this involved values for both CRL and HC obtained from the mean of three highly standardized measures of the same individual at each visit. When this protocol was applied to the INTERGROWTH-21<sup>st</sup> fetal growth data, which constituted 128 comparisons of fetal CRL and HC from early pregnancy to term, as well as birth length, only one value was marginally higher (SSD -0.58)<sup>16</sup> than the primary cut-off of  $\pm 0.5$  SD, recommended by the WHO MGRS Group.<sup>44</sup>



Now, the final key question is what the two recently published fetal studies' results would have been had the investigators followed the same analytical strategy as the WHO MGRS Group?<sup>18, 44</sup> To explore this question, we have produced the matching analyses, which the authors did not perform. Figure 1 clearly shows that, when the results of the two studies are combined, the differences amongst all the study sites are well within the limits established *a priori* for fetal skeletal measures in agreement with the INTERGROWTH-21<sup>st</sup> previous publications. This was to be expected because the actual 50<sup>th</sup> centile of fetal HC according to gestational age was almost identical across these populations (Figure 2).

### **Race/ethnicity is not a biological factor influencing fetal growth**

Several groups have suggested adjusting fetal growth charts for maternal characteristics, in particular self-reported race/ethnicity. However, the use of race/ethnicity is problematic in most non-isolated populations because of large ancestral admixture due to global migration, invasions and other population movements. There are also at least 116 definitions of self-reported race/ethnicity in the biomedical literature.<sup>46</sup>

The alternative more compelling view is that race/ethnicity is simply a social construct that represents a proxy for SES, education and social class background, which is related to many poor health and social outcomes, e.g. stillbirth and <sup>47</sup> maternal mortality <sup>48</sup> In the NICHD fetal study, which proposed using different charts in the USA for 'non-Hispanic whites, non-Hispanic blacks, Hispanic and Asian/Pacific islanders' differences seen in fetal growth between these groups must be taken in the context of the enormous differences in annual family income and other SES markers such as marital status, education and private health insurance. In addition, the fact that SES and race/ethnicity often merge is totally ignored when making adjustments and, in our view, presents a dangerous precedent.

There is also no scientific evidence that self-reported race/ethnicity is biologically or genetically related to fetal growth. Actually, all the genetic evidence across global populations demonstrates the opposite, i.e. only a very small proportion of human linear growth is related to genetic factors. As measured in observational studies, the differences between geographic locations cannot explain more than 10% of the variability in human length. In addition, strong genetic evidence from a multi-ancestry, genome-wide association study meta-analysis involving 153,781 participants, identified as many as 60 genetic loci associated with birth weight (as a proxy for fetal growth) with only 15% of the variance in birth weight being captured by assays of fetal genetic variation.<sup>49</sup>

Furthermore, in any country, such as the USA, whose inhabitants often have mixed ancestral backgrounds, it is impossible to see how racial/ethnic classification could even be implemented during the course of routine antenatal care, especially as some of the groupings are hardly scientific. For example, based on USA census practice, "Asian" includes Chinese, Japanese, South Indian and Pacific Islander groups. Black American and African Americans are presumably grouped together, although health behaviors between African-born and American-born blacks are acknowledged to be different.<sup>50</sup> Hispanics are presumably an ethnic group of European origin with (or without) native-American mixing? In fact, the infeasibility and inaccuracy of defining race/ethnicity in contemporary



multicultural settings, was recently demonstrated in an Australian study,<sup>51</sup> using the gestation-related optimal weight (GROW) customized charts.<sup>52</sup> In a sense, however, the impracticality is an irrelevance because racial/ethnic specific charts are indefensible on biological grounds.<sup>53</sup>

Interestingly, one of the quotations often used to justify having racial/ethnic fetal growth charts is from a Bogin et al. (2010) paper although the authors themselves actually provided rather different views. They stated that “even if specific genotypes are discovered, their direct contribution to normal ethnic (so-called ‘racial’) variation in human body shape may be relatively small. At 40 weeks’ gestation, fetuses identified as African-Americans have, on average, relatively longer legs than fetuses identified as European-Americans. But the difference, as measured by (total length/CRL) is less than 1%.”<sup>54</sup> Such views and those of many other scientists accord with the belief that race/ethnicity is a social rather than a biological construct and a form of categorization that is ill-defined, especially in populations that have experienced high ancestral admixture rates.

The rationale for adjusting for other factors is similarly questionable. For example, parity (nulliparous women have on average smaller fetuses, but are also at higher risk of other features of placental insufficiency such as pre-eclampsia), or maternal weight, highly dependent on over- and under-nutrition are questionable and are not unchanging characteristics. Even characteristics that do not change within an individual’s life - such as maternal height - are highly changeable within just a few generations and therefore nutritionally dependent.

### **Implications for screening in the general pregnant population: local charts versus international standards**

It is suggested that fetal growth charts for EFW and common ultrasound biometric measures after 14 weeks’ gestation “reveal a wide range of variation in human fetal growth across different parts of the world” with “significant differences in fetal growth between countries”. It is hard to understand how borders between countries, often drawn on maps by colonial powers, can possibly have a biological influence on human growth, nor how heterogeneity within populations can be negated by national boundaries.

So, what happens if “population-specific high-quality reference intervals” for each population are created? Apart from the obvious hindrance of having to create hundreds of high-quality reference intervals for countries, regions, cities, villages or hospitals, this approach is entirely fallacious. If a reference range is created for each region, by definition 10% of fetuses and newborns will be below the 10<sup>th</sup> centile of each local chart, and 10% above the 90<sup>th</sup> centile. Pretending that a uniform proportion of the population of fetuses across the world have the same degree of growth aberration is nonsensical and entirely at odds with differences in rates of maternal obesity, diabetes, pre-eclampsia, malnutrition and infectious diseases. ‘Fixing’ charts in this way would mean that no country, region or city would have an excess of underweight or overweight babies - a concept so far removed from common sense and biological principles as to be difficult to comprehend.

This has been unequivocally demonstrated in two recently published re-analyses of data from low- and middle-income countries, using a birth weight less than the 10<sup>th</sup> centile of the INTERGROWTH-21<sup>st</sup> Newborn Size for Gestational Age/Sex Standards<sup>55</sup> as the definition of SGA. In the first study, the overall SGA prevalence was 24% amongst 16 birth cohorts;<sup>56</sup> in the second, the rate was 19% amongst 14 cohorts.<sup>12</sup> Of course, the prevalence would have been 10% had local charts been used (Figure 3).

At the other end of the health spectrum, we have previously shown that in England in 2011-12 there were, as expected, 11% live singleton babies born after 33 weeks' gestation over the 90<sup>th</sup> centile if local charts of birth weight for gestational age are used.<sup>57</sup> However, when INTERGROWTH-21<sup>st</sup> international standards are used for the cut-off point,<sup>55</sup> the rate of overweight newborns increased to 19% overall, which matches the high prevalence of obesity in pregnant women and children in England (Figure 4).<sup>58, 59</sup>

An additional practical issue specific to the WHO sponsored reference charts, in terms of their global use for screening, is that they are sex-specific and their use presupposes prenatal sex determination. Even if we assume that the determination is accurate and that parents want to know the fetal sex, the practice is banned in some countries<sup>60</sup>.

## Conclusions

The WHO Child Growth Standards<sup>15</sup> are now used in nearly every country in the world to measure the growth of children from 0 to 5 years of age.<sup>61</sup> The INTERGROWTH-21<sup>st</sup> Project was designed using exactly the same prescriptive approach as the WHO MGRS, i.e. based on WHO recommendations regarding the construction of human growth standards.<sup>24</sup> The charts generated by the WHO MGRS and INTERGROWTH-21<sup>st</sup> Project integrate perfectly so that, for the first time in history, a uniform method exists for monitoring linear growth from the "womb to classroom".<sup>62</sup>

Many of the clinical tools derived from the same healthy cohort as the Fetal Growth Standards are now being used routinely around the world, e.g. the Preterm Postnatal Growth Standards, which were adopted by both WHO<sup>13</sup> and CDC<sup>14</sup> in the context of the Zika epidemic. These and other standards derived from the same cohort, e.g. for measuring symphysis-fundal height<sup>29</sup> and maternal weight gain,<sup>30</sup> as well as tools for estimating gestational age in early<sup>2</sup> and late<sup>63</sup> pregnancy, have had over 65,000 downloads from our website (data up until 21 November 2017) and close to 10,000 health care professionals have been trained using INTERGROWTH-21<sup>st</sup> e-learning modules.

There is no scientific rationale for using local references instead of standards in clinical practice, and customization based on the color of a mother's skin, the sex of her fetus or her nationality is unacceptable in the 21<sup>st</sup> century. Furthermore, classifying any of the 23.3 million infants born SGA in low- to middle-income countries according to the INTERGROWTH-21<sup>st</sup> Newborn Size Standards for Gestational Age/Sex<sup>12</sup> as 'normally grown' by local charts would deprive them of their right to better health care given that most are SGA because of impaired fetal growth due to malnutrition and/or infectious diseases.

## Panel 1: INTERGROWTH-21<sup>st</sup> Project characteristics

### **Large prospective study of 59,137 pregnant women**

**Population based:** all institutions providing pregnancy and delivery care in eight geographically limited urban areas with low rates of adverse perinatal outcomes and low pollution, domestic smoke, radiation and other toxic substances

**Sampling of individual women within the eight geographic areas** using predefined criteria for the construction of the standards

**Participants followed-up to 2 years of age**

**Pregnancy, neonatal anthropometry, and perinatal conditions recorded for the total population (59,137 pregnant women) in the eight geographic areas** using standardized procedures, identical equipment and centrally trained staff

**Environmental conditions evaluated** using a special data collection form developed in collaboration with the Centre for Environmental Research and Children's Health (CERCH), University of California, following WHO recommendations

**Excluded** from the standards **only severe maternal or fetal conditions** defined *a priori*.

**A priori data analysis plan based on WHO recommendations** to construct human growth standards

Use of **skeletal growth measures from <14 weeks' gestation to 2 years of age** for comparisons across populations, as recommended by WHO

**Three complementary data analysis strategies** to support pooling data for the construction of the standards

**International standards for human growth from <14 weeks' gestation to 2 years of age**

**International preterm postnatal growth standards** as recommended by WHO

**Preterm postnatal motor development** assessment following WHO milestones

### **Published real-time, on-line data management system**

**Ultrasound equipment selected** based on pre-defined criteria after extensive public consultation **according to WHO administrative requirements**

**Ultrasound measures in triplicate** and corroborated by **newborn anthropometry**

**Ultrasound results masked** to the operators to eliminate 'expected result bias'

**Standardized equipment** at all sites for ultrasound; maternal, newborn and child anthropometry

**Ultrasound machines calibrated with a standard phantom**

**Published system of:**

- Training, standardization and certification of ultrasound operators
- Quality control strategy for all maternal and postnatal measures
- Assessment of intra- and inter-observer variation of ultrasound fetal biometry
- Protocols for quality control of ultrasound image review, data monitoring and random sample re-measurement

**Panel 3: INTERGROWTH-21<sup>st</sup> international standards for monitoring growth and development from early pregnancy to 2 years of age**

| <b>INTERGROWTH 21<sup>st</sup> International Standards for:</b>  | <b>INTERGROWTH 21<sup>st</sup> Systematic review and Conceptual basis supporting the International Standards</b>   |
|--|--|
| First trimester dating <sup>2</sup>  | Systematic review of charts of pregnancy dating by fetal crown-rump length <sup>11</sup>   |
| Late pregnancy dating <sup>63</sup>  | Study design and implementation <sup>26</sup><br>Ultrasound methodology, standardization and quality control <sup>36, 64-68</sup><br>Systematic review of charts of fetal size by ultrasound <sup>21</sup><br>Systematic review of novel biomarkers for predicting intrauterine growth restriction and stillbirths <sup>69, 70</sup><br>Systematic review and meta-analysis on the predictive accuracy of the cerebroplacental ratio for adverse perinatal and neurodevelopmental outcomes in suspected fetal growth restriction <sup>71</sup> |
| Fetal growth by ultrasound <sup>1</sup>  |  |
| Estimated fetal weight by ultrasound <sup>3</sup>  |  |
| Symphysis-fundal height <sup>29</sup>  |  |
| Phenotypic classification of SGA <sup>72</sup><br>Phenotypic classification of preterm birth <sup>73</sup>   | Conceptual issues on preterm birth <sup>74-76</sup><br>Systematic review of novel biomarkers for the prediction of the spontaneous preterm birth phenotype <sup>77</sup>   |
| Newborn size for gestational age and sex from 24 weeks' gestation to term <sup>55</sup>  | Systematic review of charts of newborn anthropometry <sup>22</sup><br>Conceptual issues for preterm standards <sup>78</sup><br>Systematic review of preterm postnatal charts <sup>23</sup><br>Preterm postnatal growth: new paradigm <sup>79</sup><br>Anthropometric protocols, standardization and quality control methods for international growth standards <sup>80, 81</sup>   |
| Newborn body composition and weight for length standards <sup>31</sup>   |  |
| Preterm postnatal growth based on international feeding recommendations <sup>32</sup>  |  |
| Maternal weight gain during pregnancy <sup>30</sup>  | Systematic review of gestational weight gain charts <sup>82</sup>  |
| Postnatal follow-up to 2 years of age with neurodevelopmental assessment to evaluate the appropriateness of the population for creating growth standards <sup>17</sup> | Systematic review of the differential effects of intrauterine growth restriction on childhood development <sup>83</sup><br>A simplified multi-dimensional set of neurodevelopment assessment tools <sup>37</sup>   |
| Free e-learning training courses <sup>84, 85</sup>   | Evaluation of dissemination activities <sup>86, 87</sup>   |

1 **Panel 2: The different studies of the INTEGROWTH-21<sup>st</sup> project**

|  |   |
|--|---|
| Newborn Cross-Sectional Study (NCSS),                      | Demographic and pregnancy characteristics, Birth length, head circumference, weight and neonatal conditions of all newborn babies from the eight geographically defined populations using identical methods and instruments |
| Fetal growth longitudinal study (FGLS)                     | A subgroup of NCSS: Women who met individual inclusion criteria from these populations were followed up from < 9 weeks) through to the end of pregnancy;  |
| Infant follow up study (IFS)                               | All FGLS newborns were followed up until the age of two years.  |
| Preterm postnatal Follow up Study (PPFS)                   | All preterm birth in the FGLS cohort that underwent detailed regular anthropometry and followed up to the age of two years.   |
| Preterm and Impaired Fetal Growth Syndromes Study (PIFGSS) | Nested Case-Control study including all preterm birth as well as all newborns with impaired fetal growth from the NCSS  |

2

## References

1. PAPAGEORGHIU AT, OHUMA EO, ALTMAN DG, et al. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* 2014;384:869-79.
2. PAPAGEORGHIU 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 of pregnancy. *Ultrasound Obstet Gynecol* 2014;44:641-8.
3. STIRNEMANN J, VILLAR J, SALOMON LJ, et al. International estimated fetal weight standards of the INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* 2017;49:478-86.
4. KISERUD T, PIAGGIO G, CARROLI G, et al. The World Health Organization Fetal Growth Charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Med* 2017;14:e1002220.
5. BUCK LOUIS GM, GREWAL J, ALBERT PS, et al. Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. *Am J Obstet Gynecol* 2015;213:449 e1-49 e41.
6. ALBERT PS, GRANTZ KL. Fetal growth and ethnic variation. *Lancet Diabetes Endocrinol* 2014;2:773.
7. GARDOSI J. Fetal growth and ethnic variation. *Lancet Diabetes Endocrinol* 2014;2:773-4.
8. MCCARTHY EA, WALKER SP. International fetal growth standards: one size fits all. *Lancet* 2014;384:835-6.
9. STEER PJ. Possible differences in fetal size by racial origin. *Lancet Diabetes Endocrinol* 2014;2:766-7.
10. CRAMON-TAUBADEL NV. Evolutionary insights into global patterns of human cranial diversity: population history, climatic and dietary effects. *J Anthropol Sci* 2014;92:43-77.
11. NAPOLITANO R, DHAMI J, OHUMA EO, et al. Pregnancy dating by fetal crown-rump length: a systematic review of charts. *BJOG* 2014;121:556-65.
12. LEE ACC, KOZUKI N, COUSENS S, et al. Estimates of burden and consequences of infants born small for gestational age in low and middle income countries with INTERGROWTH-21st standard: analysis of CHERG datasets. *BMJ* 2017;358:j3677.
13. WHO. Screening, assessment and management of neonates and infants with complications associated with Zika virus exposure in utero: Interim guidance update, 2016.
14. CDC. Congenital microcephaly case definitions, 2016.
15. DE ONIS M, GARZA C, ONYANGO AW, MARTORELL R. WHO Child Growth Standards. *Acta Paediatr Suppl* 2006;450:1-101.
16. VILLAR J, PAPAGEORGHIU AT, PANG R, et al. The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: the Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. *Lancet Diabetes Endocrinol* 2014;2:781-92.
17. VILLAR J, CHEIKH ISMAIL L, STAINES-URIAS E, et al. The satisfactory growth and development at 2 years of age of the INTERGROWTH-21st Fetal Growth Standards cohort confirms its appropriateness for the construction of international, prescriptive, growth standards *Am J Obstet Gynecol* in press.



18. DE ONIS M, GARZA C, VICTORA CG, ONYANGO AW, FRONGILLO EA, MARTINES J. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. *Food Nutr Bull* 2004;25:S15-26.
19. DWYER-LINDGREN L, BERTOZZI-VILLA A, STUBBS RW, et al. Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA Intern Med* 2017;177:1003-11.
20. DWYER-LINDGREN L, STUBBS RW, BERTOZZI-VILLA A, et al. Variation in life expectancy and mortality by cause among neighbourhoods in King County, WA, USA, 1990–2014: a census tract-level analysis for the Global Burden of Disease Study 2015. *Lancet Public Health* 2017;2:e400-e10.
21. 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.
22. GIULIANI F, OHUMA E, SPADA E, et al. Systematic review of the methodological quality of studies designed to create neonatal anthropometric charts. *Acta Paediatr* 2015;104:987-96.
23. GIULIANI F, CHEIKH ISMAIL L, BERTINO E, et al. Monitoring postnatal growth of preterm infants: present and future. *Am J Clin Nutr* 2016;103:635S-47S.
24. WHO. An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. *Bull World Health Organ* 1995;73:165-74.
25. DE ONIS M, BRANCA F. Childhood stunting: a global perspective. *Maternal & child nutrition* 2016;12:12-26.
26. 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.
27. STEVENS GA, FINUCANE MM, PACIOREK CJ, et al. Trends in mild, moderate, and severe stunting and underweight, and progress towards MDG 1 in 141 developing countries: a systematic analysis of population representative data. *Lancet* 2012;380:824-34.
28. 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.
29. PAPAGEORGHIU AT, OHUMA EO, GRAVETT MG, et al. International standards for symphysis-fundal height based on serial measurements from the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project: prospective cohort study in eight countries. *BMJ* 2016;355:i5662.
30. CHEIKH ISMAIL L, BISHOP DC, PANG R, et al. Gestational weight gain standards based on women enrolled in the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project: a prospective longitudinal cohort study. *BMJ* 2016;352:i555.
31. VILLAR J, PUGLIA FA, FENTON TR, et al. Body composition at birth and its relationship with neonatal anthropometric ratios: the newborn body composition study of the INTERGROWTH-21st project. *Pediatr Res* 2017;82:305-16.
32. VILLAR J, GIULIANI F, BHUTTA ZA, et al. Postnatal growth standards for preterm infants: the Preterm Postnatal Follow-up Study of the INTERGROWTH-21st Project. *Lancet Glob Health* 2015;3:e681-91.
33. IOANNOU C, SARRIS I, HOCH L, et al. Standardisation of crown-rump length measurement. *BJOG* 2013;120 Suppl 2:38-41.

34. PAPAGEORGHIU 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.
35. 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 2:33-7.
36. CAVALLARO A, ASH ST, NAPOLITANO R, et al. Quality control of ultrasound for fetal biometry: results from the INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* 2017.
37. FERNANDES M, STEIN A, NEWTON CR, et al. The INTERGROWTH-21st Project Neurodevelopment Package: A Novel Method for the Multi-Dimensional Assessment of Neurodevelopment in Pre-School Age Children. *PLoS One* 2014;9:e113360.
38. GROUP WHOMGRS. WHO Motor Development Study: windows of achievement for six gross motor development milestones. *Acta Paediatr Suppl* 2006;450:86-95.
39. DE ONIS M, HABICHT J. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *Am J Clin Nutr* 1996;64:650-58.
40. HADLOCK FP, HARRIST RB, SHARMAN RS, DETER RL, PARK SK. Estimation of fetal weight with the use of head, body, and femur measurements - a prospective study. *Am J Obstet Gynecol* 1985;151:333-7.
41. OKLAND I, BJASTAD TG, JOHANSEN TF, GJESSING HK, GROTTUM P, EIK-NES SH. Narrowed beam width in newer ultrasound machines shortens measurements in the lateral direction: fetal measurement charts may be obsolete. *Ultrasound Obstet Gynecol* 2011;38:82-7.
42. DUDLEY NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 2005;25:80-9.
43. ZHANG C, HEDIGER ML, ALBERT PS, et al. Association of maternal obesity with longitudinal ultrasonographic measures of fetal growth: Findings from the NICHD fetal growth studies-singletons. *JAMA pediatrics* 2017.
44. BORCHI E, DE ONIS M, GARZA C, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Stat Med* 2006;25:247-65.
45. VILLAR J, ALTMAN DG, VICTORA CG, et al. Fetal growth and ethnic variation--authors' reply. *Lancet Diabetes Endocrinol* 2014;2:774-5.
46. MA IW, KHAN NA, KANG A, ZALUNARDO N, PALEPU A. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *J Clin Epidemiol* 2007;60:572-8.
47. SEATON SE, YADAV KD, FIELD DJ, KHUNTI K, MANKTELOW BN. Birthweight centile charts for South Asian infants born in the UK. *Neonatology* 2011;100:398-403.
48. METCALFE A, WICK J, RONKSLEY P. Racial disparities in comorbidity and severe maternal morbidity/mortality in the United States: an analysis of temporal trends. *Acta Obstet Gynecol Scand* 2017.
49. HORIKOSHI M, BEAUMONT RN, DAY FR, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature* 2016;538:248-52.
50. DAVID RJ, COLLINS JW, JR. Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites. *N Engl J Med* 1997;337:1209-14.

51. LOCKIE E, MCCARTHY EA, HUI L, CHURILOV L, WALKER SP. Feasibility of using self-reported ethnicity in pregnancy according to the gestation-related optimal weight classification: a cross-sectional study. *BJOG* 2017.
52. GARDOSI J, FIGUERAS F, CLAUSSEON B, FRANCIS A. The customised growth potential: an international research tool to study the epidemiology of fetal growth. *Paediatr Perinat Epidemiol* 2011;25:2-10.
53. Illuminating BiDiL. *Nat Biotechnol* 2005;23:903.
54. BOGIN B, VARELA-SILVA MI. Leg length, body proportion, and health: a review with a note on beauty. *International journal of environmental research and public health* 2010;7:1047-75.
55. VILLAR J, CHEIKH ISMAIL L, VICTORA CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014;384:857-68.
56. KOZUKI N, KATZ J, CHRISTIAN P, et al. Comparison of US Birth Weight References and the International Fetal and Newborn Growth Consortium for the 21st Century Standard. *JAMA pediatrics* 2015;169:e151438.
57. FREEMAN JV, COLE TJ, CHINN S, JONES PR, WHITE EM, PREECE MA. Cross sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* 1995;73:17-24.
58. HESLEHURST N, RANKIN J, WILKINSON JR, SUMMERBELL CD. A nationally representative study of maternal obesity in England, UK: trends in incidence and demographic inequalities in 619 323 births, 1989-2007. *Int J Obes (Lond)* 2010;34:420-8.
59. BOODHNA G. Children's body mass index, overweight and obesity *Health Survey for England 2013*. London, 2014 (vol 1).
60. GEORGE SM. Millions of missing girls: from fetal sexing to high technology sex selection in India. *Prenatal diagnosis* 2006;26:604-9.
61. DE ONIS M, ONYANGO A, BORCHI E, et al. Worldwide implementation of the WHO Child Growth Standards. *Public Health Nutr* 2012;15:1603-10.
62. VILLAR J, PAPAGEORGHIU AT, PANG R, et al. Monitoring human growth and development: a continuum from the womb to the classroom. *Am J Obstet Gynecol* 2015;213:494-9.
63. PAPAGEORGHIU AT, KEMP B, STONES W, et al. Ultrasound-based gestational-age estimation in late pregnancy. *Ultrasound Obstet Gynecol* 2016;48:719-26.
64. SARRIS I, IOANNOU C, DIGHE M, et al. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound Obstet Gynecol* 2011;38:681-7.
65. PAPAGEORGHIU AT, 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, v.
66. SARRIS I, IOANNOU C, OHUMA EO, et al. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. *BJOG* 2013;120 Suppl 2:33-7, v.
67. IOANNOU C, SARRIS I, HOCH L, et al. Standardisation of crown-rump length measurement. *BJOG* 2013;120 Suppl 2:38-41, v.
68. WANYONYI SZ, NAPOLITANO R, OHUMA EO, SALOMON LJ, PAPAGEORGHIU AT. Image-scoring system for crown-rump length measurement. *Ultrasound Obstet Gynecol* 2014;44:649-54.

69. CONDE-AGUDELO A, BIRD S, KENNEDY SH, VILLAR J, PAPAGEORGHIU AT. First- and second-trimester tests to predict stillbirth in unselected pregnant women: a systematic review and meta-analysis. *BJOG* 2015;122:41-55.
70. CONDE-AGUDELO A, PAPAGEORGHIU AT, KENNEDY SH, VILLAR J. Novel biomarkers for predicting intrauterine growth restriction: a systematic review and meta-analysis. *BJOG* 2013;120:681-94.
71. CONDE-AGUDELO A VJ, KENNEDY S, PAPAGEORGHIU A. . Predictive accuracy of the cerebroplacental ratio for adverse perinatal and neurodevelopmental outcomes in suspected fetal growth restriction: a systematic review and meta-analysis. Submitted 2017.
72. VICTORA CG, VILLAR J, BARROS FC, et al. Anthropometric characterization of impaired fetal growth: Risk factors for and prognosis of newborns with stunting or wasting. *JAMA pediatrics* 2015;169:e151431.
73. BARROS FC, PAPAGEORGHIU AT, VICTORA CG, et al. The distribution of clinical phenotypes of preterm birth syndrome: implications for prevention. *JAMA Pediatr* 2015;169:220-9.
74. GOLDENBERG RL, GRAVETT MG, IAMS J, et al. The preterm birth syndrome: issues to consider in creating a classification system. *Am J Obstet Gynecol* 2012;206:113-8.
75. KRAMER MS, PAPAGEORGHIU A, CULHANE J, et al. Challenges in defining and classifying the preterm birth syndrome. *Am J Obstet Gynecol* 2012;206:108-12.
76. VILLAR J, PAPAGEORGHIU AT, KNIGHT HE, et al. The preterm birth syndrome: a prototype phenotypic classification. *Am J Obstet Gynecol* 2012;206:119-23.
77. CONDE-AGUDELO A, PAPAGEORGHIU AT, KENNEDY SH, VILLAR J. Novel biomarkers for the prediction of the spontaneous preterm birth phenotype: a systematic review and meta-analysis. *BJOG* 2011;118:1042-54.
78. VILLAR J, KNIGHT HE, DE ONIS M, et al. Conceptual issues related to the construction of prescriptive standards for the evaluation of postnatal growth of preterm infants. *Arch Dis Child* 2010;95:1034-8.
79. VILLAR J GF, BARROS F, ROGGERO P, CORONADO ZARCO IA, REGO MAS, OCHIENG R, GIANNI ML, RAO S, LAMBERT A, RYUMINA I, BRITTO C, CHAWLA D, CHEIKH ISMAIL L, ALI SR, HIRST J, TEJI JS, ABAWI K, ASIBEY J, AGYEMAN-DUAH J, MCCORMICK K, BERTINO E, PAPAGEORGHIU AT, FIGUERAS-ALOY J, BHUTTA ZA, KENNEDY SH. Monitoring the postnatal growth of preterm infants: a paradigm change. . *Pediatrics* 2017;82:305-16.
80. JAFFER YA, AL ABRI J, ABDAWANI J, et al. Implementation of the INTERGROWTH-21st Project in Oman. *BJOG* 2013;120 Suppl 2:111-6, v.
81. CHEIKH ISMAIL L, KNIGHT HE, OHUMA EO, 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, v.
82. OHADIKE CO, CHEIKH-ISMAIL L, OHUMA EO, et al. Systematic Review of the Methodological Quality of Studies Aimed at Creating Gestational Weight Gain Charts. *Adv Nutr* 2016;7:313-22.
83. MURRAY E, FERNANDES M, FAZEL M, KENNEDY SH, VILLAR J, STEIN A. Differential effect of intrauterine growth restriction on childhood neurodevelopment: a systematic review. *BJOG* 2015;122:1062-72.

- 1 84. LAMPL M, KUZAWA CW, JEANTY P. Growth patterns of the heart and kidney suggest  
2 inter-organ collaboration in facultative fetal growth. American journal of human  
3 biology : the official journal of the Human Biology Council 2005;17:178-94.
- 4 85. HEMACHANDRA AH, KLEBANOFF MA. Use of serial ultrasound to identify periods of fetal  
5 growth restriction in relation to neonatal anthropometry. American journal of  
6 human biology : the official journal of the Human Biology Council 2006;18:791-7.
- 7 86. ABAWI K, GERTISER L, IDRIS R, et al. A large-scale Internet /computer-based, training  
8 module: dissemination of evidence-based management of postpartum hemorrhage  
9 to front-line health care workers. International Journal on E-Learning 2017;16:317-  
10 28.
- 11 87. CHATFIELD A, CAGLIA JM, DHILLON S, et al. Translating research into practice: the  
12 introduction of the INTERGROWTH-21st package of clinical standards, tools and  
13 guidelines into policies, programmes and services. BJOG 2013;120 Suppl 2:139-42, v.  
14  
15

## Figure legends

**Figure 1.** Standardised study discrepancy (SSD) of fitted fetal head circumference. Study-specific means were obtained as an unweighted average of values from the published charts (by gestational week) for five gestational age intervals. The SSD was calculated as the difference between the individual study mean and the mean of all studies combined, divided by the FGLS adjusted standard deviation<sup>7</sup>, at each gestational age interval. (WHO sponsored study<sup>15</sup>: grey circles; NICHD Fetal Growth Studies<sup>19</sup>: red circles (White), squares (Black), triangles (Hispanic) and diamonds (Asian); FGLS of the INTERGROWTH-21st<sup>20</sup>: green circles).

**Figure 2.** Comparison of fitted 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> centiles of fetal head circumference. WHO sponsored study<sup>15</sup>: grey solid lines; NICHD Fetal Growth Studies<sup>19</sup>: red lines (White: solid, Black: dashed, Hispanic: dash-dot, Asian: dot-dot-dash); FGLS of the INTERGROWTH-21st Project<sup>20</sup>: green solid lines.

**Figure 3.** Rates of Small for Gestational Age (SGA) from 16 prospective cohorts of newborns from 10 low- and middle-income countries. Empty columns show the prevalence of SGA using INTERGROWTH-21st standards<sup>47</sup>, compared with the effect of using a fixed cut off SGA rate of 10% that would result from using local reference charts (black columns). Data from Kozuki et al<sup>48</sup>.

**Figure 4.** Prevalence of large for gestational age (LGA, >90<sup>th</sup> birthweight centile) newborns in England in 2011-2012. The blue line corresponds to the estimated prevalence of LGA using the British 1990 growth reference centiles<sup>49</sup>; the green line corresponds to the estimated prevalence of LGA using the INTERGROWTH-21st Newborn Size at Birth Standards<sup>47</sup>.