Implementing the INTERGROWTH-21st fetal growth standards in France: a ‘flash study’ of the College Français d’Echographie Foetale (CFEF)

J. J. STIRNEMANN*†, N. FRIES‡, R. BESSIS‡, M. FONTANGES‡, R. MANGIONE‡ and L. J. SALOMON*†‡

*Assistance Publique-Hôpitaux de Paris, Hôpital Necker-Enfants Malades, Paris, France; †EA FETUS, 7328, Université Paris-Descartes, Paris, France; ‡Collège Français d’Echographie Foetale, CFEF, France

KEYWORDS: biometry; congenital abnormalities; cross-sectional studies; epidemiology; fetal development; growth charts; pregnancy; ultrasound

ABSTRACT

Objectives To assess potential differences in fetal size between the French population and the international population from the INTERGROWTH-21st (IG-21st) Project and to measure the impact of switching to the IG-21st reference standards for fetal size.

Methods This was a nationwide cross-sectional study of fetal ultrasound biometry. Low-risk singleton pregnancies were recruited prospectively within the network of the national French College of Fetal Ultrasound, CFEF, over a 6-week period. Further selection was performed based on the criteria of the IG-21st Project in order to obtain a comparable population. Head circumference (HC) was used as the main fat-free skeletal measure of growth for comparison of French fetal size with that of the IG-21st population. The impact of switching to the IG-21st fetal growth standards was quantified by comparing Z-scores calculated using the IG-21st standards with those calculated using locally derived reference ranges for HC, abdominal circumference (AC) and femur length (FL).

Results Following selection, 4858 cases were analyzed. The distribution of HC demonstrated clear similarity between our French population and the IG-21st population: our observed centile curves closely matched those of IG-21st and the Z-scores were close to 0 across gestational age. The IG-21st standards performed as well as did locally derived charts in terms of screening for small-for-gestational age by AC, while they identified significantly fewer small FL values than were expected and than did the locally derived charts.

Conclusions Under strict selection criteria, fetal size in France is similar to that of the international population used in the IG-21st Project. The discrepancies in FL are unlikely to impact on prenatal management. Therefore, switching from locally derived reference ranges to the IG-21st standards appears to be a safe option. Copyright © 2016 ISUOG. Published by John Wiley & Sons Ltd.

INTRODUCTION

Screening for intrauterine growth restriction (IUGR), mainly by ultrasound measurements, is one of the most important components of antenatal care. Unfortunately, the antenatal detection rate of IUGR confirmed at birth is low, even in high-risk subpopulations.

The low detection rate can be explained in part by the choice of growth chart, from amongst the more than 80 available, on which to base judgments about the fetal measurements and the lack of international consensus regarding the diagnostic criteria for IUGR. This is in contrast with the pediatric evaluation of growth in infants and children, as the World Health Organization (WHO) has produced a single set of international growth standards for use worldwide.

Conceptually similar to the WHO Child Growth Standards, the International Fetal Growth Standards were published in 2014 as part of the INTERGROWTH-21st Project. These standards were developed from multiethnic populations worldwide, whose health, nutrition and care needs were largely met. As opposed to previous locally produced references, international standards have the potential to improve the detection of growth disturbances and, as a result, perinatal outcomes, by standardizing the diagnostic approach to IUGR and macrosomia.
The present cross-sectional study was undertaken to: (1) evaluate whether fetal biometry measurements from pregnant French women, using inclusion criteria similar to those of the INTERGROWTH-21st Project, produce patterns comparable to the INTERGROWTH-21st standards and (2) quantify the implications of using the INTERGROWTH-21st standards in the French population.

METHODS

Flash studies are pragmatic, short and very focused studies, conducted without modifying routine clinical practice and at no extra cost. They have both a scientific and an educational purpose and are conducted in France across the countrywide network of sonographers who are members of the French College of Fetal Ultrasound (College Français d’Echographie Foetale (CFEF)).

We invited sonographers first to take an online training course (www.cef.fr) that reviewed the aims of the study, the inclusion criteria, the methodology for taking the measurements and the biometric quality control criteria. Only sonographers who completed the course and passed the final test were eligible to participate in the study.

Pregnant women contributed with a single measure and were included prospectively and consecutively over a fixed study period of 6 weeks. Those included had a singleton pregnancy without congenital malformations and were scanned between 8 September 2014 and 18 October 2014, based on a routine indication: in France, three scans are planned for all pregnancies, at 11–14 weeks, 21–24 weeks and 30–34 weeks, as part of standard antenatal care. Pregnancy dating was based on crown–rump length measurement in the first trimester, as recommended by the French College of Obstetrics and Gynaecology, CNGOF.

The following information was collected: parity, gravidity, maternal age, body mass index (BMI), socioprofessional category, presence of comorbidities, smoking, alcohol consumption, gestational age at scan and fetal head circumference (HC), abdominal circumference (AC) and femur length (FL). Fetal measurements were taken as described elsewhere. Briefly, HC was measured in an axial plane of the fetal head at the level at which the continuous midline echo is broken by the septum pellucidum in the anterior third, using the ellipse tool with calipers placed on the outer edges of the skull. AC was measured in an axial plane of the fetal abdomen, just above the level of the cord insertion, using the ellipse tool. FL was measured on a plane showing the entire femoral diaphysis, with both ends clearly visible and at an angle < 45° to the horizontal. From 30 weeks’ gestation onwards, particular care was taken not to include the epiphysis.

These measurements, collected prospectively, constituted our primary database. Within this dataset, a subsample of women was selected who met, as closely as possible, the strict inclusion criteria of the Fetal Growth Longitudinal Study (FGLS) of the INTERGROWTH-21st Project. Briefly, the FGLS participants were selected first at the level of a geographic area and then at the individual level within each study site. The principal geographic-level criteria were met because all deliveries in France occur in health institutions, all of which are located at an altitude of 1600 m or lower. For the individual-level criteria, similar to the FGLS, we included women who were at low risk of adverse maternal outcome (e.g. absence of smoking, diabetes, hypertension, renal disease, maternal symptoms of pre-eclampsia or other medical conditions associated with fetal growth disturbances) and perinatal outcome (patients not referred for suspicion of fetal anomaly, history of fetal anomaly or IUGR). In the final analysis we included only women aged ≥ 18 years and ≤ 35 years, who had BMI ≥ 18.5 kg/m² and < 30 kg/m², who had no clinically relevant obstetric or gynecological history and who initiated antenatal care before 14 weeks’ gestation. We also excluded a-priori measurements below or above −5 and +5 SD, respectively.

Our data were compared to the INTERGROWTH-21st standards and a method similar to that described in the original studies that generated the WHO Child Growth and INTERGROWTH-21st standards. To assess the similarity of size between French fetuses and those used to create the INTERGROWTH-21st standards, we decided to use HC, the main fat-free skeletal measure of linear growth, for comparison. First, the charts obtained from fitting an ‘INTERGROWTH-like’ fractional polynomial regression model (i.e. the same functional form defined by the set of fractional polynomial powers) to our newly collected data were compared visually to those from the INTERGROWTH-21st standards. Similarities in the French and INTERGROWTH-21st populations with respect to the fetal HC were identified based on standardized differences, at various gestational-age windows. Standardized differences were defined as the difference between the mean of our French sample and the mean of the pooled INTERGROWTH-21st data, expressed as units of the pooled SD observed in the corresponding INTERGROWTH-21st study. Standardized differences were therefore similar to a Z-score. A standardized difference value < 0.5 was prespecified as reflecting adequately that the French population could have been included in the pooled dataset used to construct the international standards.

We also evaluated the impact of switching from the locally derived reference charts that are presently recommended for use in France to the INTERGROWTH-21st standards by calculating the mean and SD of computed Z-scores based on these two different references. A mean Z-score < 0.5 was again considered to demonstrate good concordance between observed and reference-based predicted values. A SD of Z-scores close to 1 (i.e. between 0.8 and 1.2) was considered to reflect good concordance between observed and reference-based predicted dispersion of values. Finally, we calculated the proportions of fetuses below the 3rd and 10th centiles and above the 90th and 97th centiles, using both the INTERGROWTH-21st and the
RESULTS

There were 160 sonographers who agreed to participate, of whom 120 fulfilled the inclusion criteria of the study. During the study period they performed a total of 8784 scans, of which 1689 were excluded for at least one of the following reasons: maternal condition \( (n = 1090) \); maternal symptoms \( (n = 311) \); history of previous small baby, suspected fetal anomaly or IUGR \( (n = 599) \). A further 1248 \( (14.2\%) \) were excluded because maternal age was \( > 35 \) years \( (n = 1220) \) or \( < 18 \) years \( (n = 28) \), and 961 \( (10.9\%) \) were excluded because BMI was \( < 18.5 \mathrm{kg/m}^2 \) \( (n = 193) \) or \( > 30 \mathrm{kg/m}^2 \) \( (n = 768) \). Lastly, 28 examinations were excluded because measurement values were below or above \(-5\) or \(+5\) SD, respectively.

From the original 8784 scans, our analysis thus included data from 4858 \( (55.3\%) \) independent ultrasound examinations in low-risk women and fetuses across gestation, i.e. an ‘INTERGROWTH-21st’ FGLS-like population. Figure 1 presents the INTERGROWTH-21st standard for HC, the main, fat-free, skeletal measure of linear growth used in the INTERGROWTH-21st Project for comparisons across populations\(^7\), superimposed on the individual HC values for our French subpopulation. The centiles estimated from our French subpopulation using the same statistical methodology as that of the FGLS (i.e. same set of fractional polynomial powers) are also superimposed. Visual inspection of the two sets of centiles demonstrates clear similarity between the two populations, suggesting that fetal HC size in the French population is comparable to that observed in the FGLS. This was further confirmed by analyzing the standardized differences of HC, which remained close to 0 and within \( \pm 0.5 \) when computed at different gestational-age windows across pregnancy (Figure 2).

To assess the impact of switching from local charts to the INTERGROWTH-21st charts, we also estimated the individual Z-scores for AC, FL and HC in our sample of the French population based on each of the two sets of charts\(^7\). The mean Z-scores remained between \(-0.5\) and \(+0.5\) for all measurements at all gestational-age windows for both the French locally-derived charts and the INTERGROWTH-21st charts (Table 1) in all except two of the 15 comparisons for each of the two charts.

In any distribution of Z-scores the expected SD is 1; we therefore compared the actual SDs obtained using the INTERGROWTH-21st charts and local charts. The INTERGROWTH-21st charts gave a SD of 1.01 for HC, 0.99 for AC and 1.00 for FL, and the local charts gave SDs of 0.75, 0.75 and 0.78, respectively; thus, the INTERGROWTH-21st charts were closer to the expected population distribution than were the local charts. This was further confirmed by the fact that two SD values out of 15 comparisons were beyond the prespecified cut-off for the INTERGROWTH-21st charts vs 12 values using the locally derived charts (Table 1). This resulted in different proportions of fetuses below the 3rd and 10th centiles and above the 90th and 97th centiles, using each chart (Table 2). Specifically, the INTERGROWTH-21st standard identified a proportion of HCs below the 3rd and 10th centiles closer to the expected value than did the locally derived charts. Despite the differences in the observed mean and SD of the distribution of Z-scores according to each chart, the proportions of ACs below the 3rd and 10th centiles according to each chart were
Table 1  Z-scores for study population of 4858 low-risk singleton French pregnancies estimated based on French fetal size charts (local)\(^{17}\) and on INTERGROWTH-21st international growth charts (IG-21st)\(^{7}\)

<table>
<thead>
<tr>
<th>Gestational age</th>
<th>n</th>
<th>Mean local</th>
<th>Mean IG-21st</th>
<th>SD local</th>
<th>SD IG-21st</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 to 18 weeks</td>
<td>196</td>
<td>-0.42</td>
<td>-0.67(^{*})</td>
<td>0.13</td>
<td>0.46</td>
</tr>
<tr>
<td>19 to 23 weeks</td>
<td>2123</td>
<td>0.08</td>
<td>-0.27</td>
<td>0.33</td>
<td>0.48</td>
</tr>
<tr>
<td>24 to 28 weeks</td>
<td>333</td>
<td>0.19</td>
<td>-0.26</td>
<td>0.03</td>
<td>0.43</td>
</tr>
<tr>
<td>29 to 33 weeks</td>
<td>1826</td>
<td>0.33</td>
<td>-0.36</td>
<td>0.12</td>
<td>0.53(^{*})</td>
</tr>
<tr>
<td>34 to 40 + weeks</td>
<td>380</td>
<td>0.38</td>
<td>-0.43</td>
<td>0.12</td>
<td>0.44</td>
</tr>
</tbody>
</table>

*Mean values below −0.5 or above +0.5 and SDs below 0.8 or above 1.2 are indicated. AC, abdominal circumference; FL, femur length; HC, head circumference.

Table 2  Percentages of 4858 low-risk French singleton pregnancies with fetus under 3rd or 10th centile or above 90th or 97th centile using international INTERGROWTH-21st prescriptive charts (IG-21st)\(^{7}\) and using locally derived French charts (local)\(^{17}\)

<table>
<thead>
<tr>
<th>IG-21st</th>
<th>&lt; 3rd centile (%)</th>
<th>&lt; 10th centile (%)</th>
<th>&gt; 90th centile (%)</th>
<th>&gt; 97th centile (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>Local</td>
<td>IG-21st</td>
<td>Local</td>
<td>IG-21st</td>
</tr>
<tr>
<td>3.71</td>
<td>0.70</td>
<td>11.20</td>
<td>2.28</td>
<td>7.95</td>
</tr>
<tr>
<td>AC</td>
<td>1.17</td>
<td>0.89</td>
<td>4.06</td>
<td>3.36</td>
</tr>
<tr>
<td>FL</td>
<td>1.03</td>
<td>2.08</td>
<td>3.13</td>
<td>9.18</td>
</tr>
<tr>
<td>Local</td>
<td>6.48</td>
<td>5.13</td>
<td>18.46</td>
<td>6.81</td>
</tr>
<tr>
<td></td>
<td>2.37</td>
<td>9.18</td>
<td>19.06</td>
<td>7.43</td>
</tr>
</tbody>
</table>

AC, abdominal circumference; FL, femur length; HC, head circumference.

DISCUSSION

This study shows that, in a French population selected using similar criteria to those of the FGLS at the individual level, HC measures closely matched the INTERGROWTH-21st standard. Indeed, the patterns of measurements obtained across gestation were very similar (Figure 1). This suggests that French fetuses from low-risk pregnancies do not differ from those included in the INTERGROWTH-21st Project.

Our results suggest that the routine implementation in France of such standards, based on a normative approach and international sampling from geographically and ethnically diverse populations, would be legitimate as a first level of screening. Furthermore, considering the multiethnic characteristics of present-day populations in the country, it is the most practical tool to be used during routine antenatal care. This also confirms the international nature of skeletal fetal growth, when health, nutrition and care needs are met, as is the case for the general French population. Finally, the experience gained in the implementation, in more than 140 countries, of the postnatal WHO Child Growth Standards, whose design and methods were almost identical to those of the INTERGROWTH-21st Project, also favors the widespread use of the new INTERGROWTH-21st standards\(^{7}\).

Our study shows that both the international standards and the locally derived charts appropriately describe mean fetal size, as mean Z-scores were close to zero in all cases. However, there were differences in the mean Z-score values between the two sets of charts and, consequently, discrepancies in the proportions of small and large measurements, which could be explained by several factors. A possible effect due to ethnic differences or differences in population background is still a matter of debate\(^{19,20}\), but is very unlikely since the INTERGROWTH-21st cohort includes countries such as Italy and the UK, with demographic characteristics very close to those of France. However, socioeconomic status has been shown to be a main determinant of growth\(^{21,22}\) and this variable was not considered at an individual level in our population. The remaining discrepancies could be attributed to ultrasound measurement methods. This is particularly relevant for the measurement of HC, which was previously obtained with calipers placed on the middle of the skull bone\(^{17}\), whereas recent recommendations\(^{12}\) suggest caliper placement on the outer edge, as in the new INTERGROWTH-21st international standard\(^{7}\).

The greatest discrepancies were found for FL: compared with the locally derived reference, the INTERGROWTH-21st chart ‘underestimated’ the proportion of small femurs. In practice, this would have very little, if any, impact: it should not affect our ability to detect skeletal dysplasias, will not affect screening for IUGR, which relies on AC, HC or estimated fetal weight,
and has long been abandoned for screening for Down syndrome.

Another important factor that may explain the observed differences is the recent technological changes in ultrasound machines. This is particularly relevant for FL measurements\(^2\), because the thinning of the modern ultrasound beam leads to shorter measurements\(^2\). This may explain the lower mean FL values in the INTERGROWTH-21\(^{st}\) standard\(^7\) compared with older references constructed using data obtained with outdated equipment\(^7\).

Finally, all sonographers involved in the flash study were accustomed to working with the former French references (expected-value bias), and these same references were displayed on the screen of their ultrasound machine at the time at which measurements were being performed (in the INTERGROWTH-21\(^{st}\) Project, sonographers were blinded to the measurements). It is therefore likely that this tended to bias measurements towards the values ‘expected’ based on the locally produced charts, possibly explaining the trend for the means of Z-scores to be closer to zero with the locally produced charts than with the new international standards.

The adequacy of the observed and expected SDs is critical and our results suggest better performance when using the new INTERGROWTH-21\(^{st}\) standards. The locally produced reference ranges currently in use in France\(^7\) are descriptive, i.e. they were derived from the general population, with no growth-related or outcome-related exclusion criteria, therefore capturing both normally and abnormally grown fetuses. The ability of these locally derived references to identify abnormal biometric measurements may therefore be hampered, reflected in the reported limited ability to pick up true IUGR cases in France\(^8\). Conversely, the new international prescriptive standards for fetal size\(^2\), which defines the expected fetal size under optimal conditions, have smaller dispersion (i.e. a SD closer to the theoretical value of 1) than do locally produced references and thus could improve the sensitivity of IUGR screening.

Limitations of this study, however, should be considered. The sonographers who volunteered and were eventually enrolled into this study may not fully represent the general population of sonographers. They performed biometric measurements and compared them to existing local references in a non-blinded fashion, which may have introduced a bias towards the expected values. Therefore, given the setting of our study, under real-life non-experimental conditions, we did not strictly reproduce the experimental setting of the INTERGROWTH-21\(^{st}\) study. Using routine cross-sectional ultrasound data instead of longitudinal follow-up data may also have affected adversely our assessment of the INTERGROWTH-21\(^{st}\) charts: (1) between the routine time-points, data were scarce; (2) since the study was not longitudinal, drop-outs were not recognized, which could have biased our population if such drop-outs were not random but related to growth disorders; (3) as this study aimed to measure prenatal biometry adequacy, we did not collect perinatal outcomes.

In conclusion, our study demonstrates that a French pregnant subpopulation, selected using individual criteria closely matching those used in the INTERGROWTH-21\(^{st}\) Project, produces fetal size patterns comparable to the INTERGROWTH-21\(^{st}\) international standards in terms of skeletal ultrasound parameters. We suggest that these international standards can be used in France as part of routine ultrasound practice. However, it is of paramount importance to separate the choice of a standard based on data quality and reference population selection from any clinical screening procedures based on centile cut-offs. Therefore, the implementation of such standards deserves close monitoring: since they were derived from a highly selected population, they are likely to increase detection rates of growth abnormalities in the general population for a given centile cut-off. Therefore, biometry-based screening procedures should be adapted accordingly, based on the future appraisal of the implementation of these charts in the general population, in terms of detection rates and perinatal outcome.

REFERENCES


