Validation of Reference Charts for Mid-Trimester Fetal Biometry

K. van de Kamp    E. Pajkrt    A.H. Zwinderman    J.A. van der Post    R.J.M. Snijders
Department of Obstetrics, University Medical Centre Amsterdam, Amsterdam, The Netherlands

Keywords
Charts · Fetal biometry · Validation · Reference · Mid-trimester

Abstract

Objective: The aim of our study was to assess charts proposed for international use in the Intergrowth-21st Project.

Methods: Ultrasound data were collected from 43,923 healthy singleton pregnancies examined at 18–23 weeks of gestation in the Netherlands. Fetal measurements were converted into Z-scores using previous and current Dutch reference charts and Intergrowth charts. The distributions of the Z-scores were compared with the expected standard normal distribution.

Results: In the Dutch population, Intergrowth curves perform well for head circumference and biparietal diameter, but not for abdominal circumference (AC, Z-score = 0.43) and femur length (FL, Z-score = 0.26). Similar findings have been reported in other European countries. Compared with the population in the Intergrowth study, Dutch women are relatively tall (170 vs. 162 cm) and sturdy (67 vs. 61 kg) with a moderately high BMI. Maternal size, in particular maternal height, is positively correlated with birthweight.

Conclusions: Whilst the establishment of the Intergrowth charts is an important step towards worldwide uniformity, especially for AC and FL. Results from our validation study indicate that distinction between normal and pathologically small babies may be improved by taking maternal size into account.

Introduction

Prenatal care is increasingly focused on the assessment of patient-specific risks to ensure that each pregnant woman receives the care that is needed whilst unnecessary interventions are avoided. To calculate individual risks, prognostic models have been developed for various maternal and fetal outcomes [1]. Fetal size is an important predictor in several models, since both small and large for gestational age fetuses are at increased risk of perinatal mortality and morbidity [2, 3].

To accurately assess fetal size, a reliable estimate of gestational age is crucial. In the Netherlands the method for pregnancy dating was standardized in 2010. The Dutch protocol prescribes that all pregnancies are dated based on crown-rump length (CRL) between 10 and 13 weeks of gestation. At the same time, clear instructions were provided to sonographers on measurement technique and caliper placement for routine fetal biometry.
Observed measurements are compared with the mean or median of a reference chart. In the Netherlands, the measurement is considered “small” if it falls under the 5th centile of the population distribution, it is classified “large” if it is above 95th centile, and it is regarded as “normal” if it is somewhere in between.

Until the year 2008, all Dutch ultrasound centers selected their own set of reference charts. The most commonly used were the charts developed in the UK by Chitty et al. [4–6] and Snijders and Nicolaides [7]. When in 2008 new charts were derived based on a study in the southwest of the Netherlands, it was decided to implement these charts as the new national gold standard [8]. However, the charts have not been validated since. The recent proposal of Papageorghiou et al. [9] to standardize assessment of fetal biometry on an international level stimulated us to evaluate the performance of the charts currently in use and to examine whether the international curves would be appropriate.

The aim of the present study was to examine the validity of the current Dutch reference charts [8] and to compare the results with previously used growth charts [4–7] and the recently proposed charts based on the Intergrowth-21st Project [9].

Material and Methods

Data Collection
A prospective cross-sectional cohort study was undertaken in the northwest region of the Netherlands. Ultrasound data were collected at 18–23 weeks of gestation in singleton pregnancies with an estimated due date between January 1, 2009, and December 31, 2012. The scans were performed in healthcare centers affiliated with the Academic Medical Centre in Amsterdam. Participants have given their informed consent, and the study protocol has been approved by the institute’s committee on human research.

All sonographers were trained to take fetal measurements according to the standard protocol developed by the Dutch Society of Obstetrics and Gynaecology (NVOG Protocol Foetale Biometrie [10]). Biparietal diameter (BPD) and head circumference (HC) were measured in a transverse section of the head with a central midline echo, interrupted in the anterior third by the cavity of the septum pellucidum with the anterior and posterior horns of the lateral ventricles in view. For BPD, the outer–outer diameter was measured perpendicular to the midline, and for HC an ellipse was projected an ellipse around the fetal abdomen, in which skin and subcutaneous tissue were within the ellipse. Femur length (FL) was measured by visualizing the diaphysis over the full length. In particular, at a more advanced pregnancy duration, the epiphysis may also have been visible, but was not involved in the measurement.

All pregnancies were dated between 10 + 0 and 12 + 6 weeks based on CRL measurement using the Robinson formula: gestational age in weeks = 8.052 × \( \sqrt{\text{CRL} \times 1.037} \) + 23.73 [11].

Data on the outcome of the pregnancies were collected from the ultrasound and birth database of the Academic Medical Center, supplemented with outcome data from other hospitals in the region and from the Dutch Perinatal Registry (Perinéd).

Exclusion Criteria
Exclusion criteria were multiple pregnancy, suspected major fetal anomaly, termination of pregnancy, miscarriage, perinatal death, and unknown outcome. In addition, records were excluded if data on either HC, AC, TCD or FL were missing. If only BPD was missing, the record remained in the study since it was an optional measurement during the study period.

Reference Charts Examined
The Chitty reference charts [4–6] were based on data from a prospective cross-sectional study of approximately 650 low- and high-risk pregnancies where ultrasound and menstrual gestational ages at 18–22 weeks differed less than 10 days. The Nicolaides charts [7] were based on a retrospective cross-sectional study of measurements from 1,040 pregnant women with a known last menstrual period (LMP) and a cycle length of 26–30 days. The Verburg charts [8] originated from a prospective longitudinal cohort study of 8,313 pregnancies that were dated based on CRL at 10–12 weeks of gestation. The Intergrowth charts [9] were derived based on the prospective longitudinal study of 4,321 pregnancies in 8 countries. LMP was used to calculate gestational age, provided that women reported a regular menstrual cycle of 24–32 days and that the discrepancy with first trimester ultrasound dates was 7 days or less. Our study focused on the measurements taken between 18 and 24 weeks of gestation.

Comparison with Intergrowth Curves
A literature search was done to identify studies examining validity of the Intergrowth curves. For each of the studies, the mean and standard deviation (SD) of the Z-scores were compared. Possible differences in baseline population characteristics including maternal weight and height, birthweight, and prevalence of diabetes and obesity were examined using data presented in the paper or national health statistics.

Statistical Analyses
For evaluation of the reference curves, all fetal measurements were transformed into Z-scores [4–9]. Z-scores were calculated using the formula: \( Z\text{-score} = \frac{X\text{GA} – M\text{GA}}{SD\text{GA}} \), where XGA was the actual fetal measurement at a given gestation and MGA and SDGA were the expected mean and SD according to the reference chart. Normality of the Z-score distribution was examined using a continuous Kolmogorov-Smirnov one-sample test (Kolmogorov-Smirnov d value). In view of the large sample sizes, a statistically significant non-normality was accepted unless the normal plot showed a clear deviation from a straight line. Mean and SD of the Z-scores were computed. Means of Z-scores were tested against...
the expected value of 0 using a t test for single samples. SDs were tested against the expected value of 1 based on the χ² distribution. The reference curves with Z-scores closest to 0 and SDs closest to 1 were considered most appropriate for our population. Moreover, the mean Z-scores were plotted against gestational age to examine whether the mean was close to zero throughout the time period for the mid-trimester scan.

To examine the impact of using different reference charts on referral for further assessment, the number of fetuses with HC, AC, and FL measurements below the 5th and above the 95th percentiles were determined (Z-score <–1.64 and >+1.64, respectively). The χ² test was used to compare the percentage of fetuses with measurements below the 5th and above the 95th percentiles with the expected percentage.

Results

Evaluation of the Reference Curves

During the study period, fetal biometry was obtained at 18–24 weeks of gestation in 55,884 consecutive pregnancies. A total of 11,961 (21.4%) cases were excluded because of multiple pregnancy (n = 1.499, 2.7%), suspected major anomaly (n = 2.024, 3.6%), termination of pregnancy (n = 12, 0.02%), miscarriage or perinatal death (n = 301, 0.5%), incomplete data (n = 2.514, 4.5%) or missing outcome of the pregnancy (n = 5.611, 10.0%). Statistical analysis was performed on 43,923 HC, AC, FL and TCD, and 30,300 BPD measurements.

None of the Z-score distributions was normally distributed, with the Kolmogorov-Smirnov d value ranging from 0.011 to 0.040 (p < 0.001). Given the large sample size, non-normality was accepted as Q-Q plots showed no clear deviation from a straight line. The Z-score mean values were between –0.518 and 0.431, and were all significantly different from zero, with the exception of the mean of Z-scores obtained for TCD measurements using the Verburg equation. The SD of the Z-scores ranged between 0.358 and 0.948; all were significantly smaller than 1.

Head Circumference

For HC, the mean Z-score obtained with the Intergrowth curve corresponded best with the expected score of 0.0 but the SD was with 0.8 lower than the expected 1.0 (Fig. 1). Using the Verburg curve, the mean Z-score decreased with gestational age from 0.3 at 18 to –0.2 at 24 weeks; the SD was close to the expected 1.0. The Nicolaides curve provided relatively high Z-scores, and the Chitty curve yielded relatively low Z-scores throughout the gestational age range.

Abdominal Circumference

For AC, the mean Z-score obtained with the Chitty curve corresponded best with the expected score of 0.0 (Fig. 2). With the Verburg curve, the mean Z-score was relatively high at 18 and 19 weeks. The Nicolaides and Intergrowth curves provided relatively high Z-scores, throughout the gestational age range. The SD of the Z-scores was closest to 1.0 using the equations of Intergrowth or Verburg.

Femur Length

For FL, the mean Z-score obtained with the Verburg curve, although relatively high at 18 and 19 weeks, corresponded best with the expected score of 0.0 (Fig. 3). The Nicolaides and the Intergrowth curve provided high Z-scores, and the Chitty curve yielded low Z-scores, throughout the gestational age range. The SD of the Z-scores was closest to 1.0 using the equation of Intergrowth.
**Biparietal Diameter**
For BPD, the mean Z-score obtained with both the Verburg and the Intergrowth curve corresponded well with the expected score of 0.0 (Fig. 4). The Nicolaides curve provided relatively high Z-scores, and the Chitty curve yielded relatively low Z-scores, throughout the gestational age range. The SD of the Z-scores was closest to 1.0 using the equation of Verburg.

**Transverse Cerebellar Diameter**
The mean Z-scores calculated with the equation from both Verburg and Nicolaides were close to 0.0, except at...
18 weeks of gestation where the mean obtained with the reference curve of Verburg was farther from 0.0 (Fig. 5). The SD of the Z-scores was closest to 1.0 with the use of the equation of Verburg.

In Table 1, an overview is presented of the number of fetuses with a Z-score below –1.64 and above 1.64 per parameter and HC, AC, and FL combined. The percentage below –1.64 per parameter ranged from 0.0 to 4.1, the percentage above 1.64 ranged from 0.2 to 8.7. All percentages were significantly different (p < 0.05) from the expected percentage. The percentage of fetuses with a Z-score of either HC, AC, or FL below –1.64 ranged from 0.1 to 6.4, and above 1.64 from 2.1 to 14.1.

Table 2 shows the mean maternal height and weight, the mean birthweight and the percentage of obesity and diabetes among adult females in countries that contrib-

**Fig. 5.** Illustration of the mean (a) and SD (b) of the Z-score of TCD measurement throughout gestation for the two different reference equations.

**Table 1.** Measurements below the 5th (Z-score <–1.64) or above the 95th percentile (Z-score >+1.64) for each parameter alone and combined using the reference curves, tested with the $\chi^2$ test against the expected percentage of 5%

<table>
<thead>
<tr>
<th>Parameter</th>
<th>First author [Ref.]</th>
<th>Z-score &lt;–1.64 n (%)</th>
<th>$\chi^2$</th>
<th>Z-score &gt;+1.64 n (%)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>Verburg [8]</td>
<td>1,034 (2.4)</td>
<td>649.1*</td>
<td>2,373 (5.4)</td>
<td>14.7*</td>
</tr>
<tr>
<td></td>
<td>Intergrowth [9]</td>
<td>421 (1.0)</td>
<td>1,511.2*</td>
<td>755 (1.7)</td>
<td>996.2*</td>
</tr>
<tr>
<td></td>
<td>Nicolaides [7]</td>
<td>3 (0.0)</td>
<td>2,306.3*</td>
<td>373 (0.8)</td>
<td>1,595.8*</td>
</tr>
<tr>
<td></td>
<td>Chitty [4–6]</td>
<td>1,799 (4.1)</td>
<td>61.4*</td>
<td>42 (0.1)</td>
<td>2,223.0*</td>
</tr>
<tr>
<td>AC</td>
<td>Verburg [8]</td>
<td>536 (1.2)</td>
<td>1,323.4*</td>
<td>2,026 (4.6)</td>
<td>14.0*</td>
</tr>
<tr>
<td></td>
<td>Intergrowth [9]</td>
<td>281 (0.6)</td>
<td>1,760.7*</td>
<td>3,832 (8.7)</td>
<td>1,280.9*</td>
</tr>
<tr>
<td></td>
<td>Nicolaides [7]</td>
<td>0 (0.0)</td>
<td>–</td>
<td>71 (0.2)</td>
<td>2,173.7*</td>
</tr>
<tr>
<td></td>
<td>Chitty [4–6]</td>
<td>858 (2.0)</td>
<td>857.7*</td>
<td>571 (1.3)</td>
<td>1,260.5*</td>
</tr>
<tr>
<td>FL</td>
<td>Verburg [8]</td>
<td>752 (1.7)</td>
<td>1,000.4*</td>
<td>1,476 (3.4)</td>
<td>249.0*</td>
</tr>
<tr>
<td></td>
<td>Intergrowth [9]</td>
<td>402 (0.9)</td>
<td>1,543.7*</td>
<td>2,460 (5.6)</td>
<td>33.2*</td>
</tr>
<tr>
<td></td>
<td>Nicolaides [7]</td>
<td>22 (0.1)</td>
<td>2,266.5*</td>
<td>754 (1.7)</td>
<td>997.6*</td>
</tr>
<tr>
<td></td>
<td>Chitty [4–6]</td>
<td>446 (1.0)</td>
<td>1,469.0*</td>
<td>342 (0.8)</td>
<td>1,648.7*</td>
</tr>
<tr>
<td>HC or AC or FL</td>
<td>Verburg [8]</td>
<td>2,093 (4.8)</td>
<td>–</td>
<td>5,044 (11.5)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Intergrowth [9]</td>
<td>1,015 (2.3)</td>
<td>–</td>
<td>6,208 (14.1)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Nicolaides [7]</td>
<td>23 (0.1)</td>
<td>–</td>
<td>1,132 (2.6)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Chitty [4–6]</td>
<td>2,817 (6.4)</td>
<td>–</td>
<td>924 (2.1)</td>
<td>–</td>
</tr>
</tbody>
</table>

HC, head circumference; AC, abdominal circumference; FL, femur length. * $p < 0.05$. 

HC, head circumference; AC, abdominal circumference; FL, femur length. * $p < 0.05$.
uted to the studies in this paper. The maternal height ranged from 159 cm in the Chinese to 170 cm in the Dutch population and the maternal weight ranged from 54 kg in the Chinese to 67 kg in the Greek and Dutch populations. There are marked differences between countries in the prevalence of mild obesity with the UK and Greece at the high end and India at the low end. Diabetes is relatively common in Brazil and the USA, whilst it is low in Kenya. The mean birthweight of the children is positively associated with height and weight of the mother ($R^2 = 0.767$ and $R^2 = 0.604$, respectively) and with the percentage of moderate and severe obesity ($R^2 = 0.548$ and $R^2 = 0.323$, respectively). Mean birthweight is negatively associated with the percentage of women with diabetes ($R^2 = -0.378$).

In Table 3, the $Z$-scores and SD obtained with the Intergrowth curves are summarized for the French, Chinese and Dutch populations. In all studies, mean $Z$-scores were within the $-0.5$ to $0.5$ range, which has been proposed as a cut-off for good concordance. Moreover, all but one of the SDs were within the proposed range of $0.8–1.2$ [12].

**Discussion**

This study aimed to determine whether the curves proposed in the Intergrowth project are appropriate for assessment of fetal biometry in the Netherlands. Our findings indicate that for the time being our population-based standards should be maintained, as with the intergrowth curves.
several cases of growth restriction may be missed. The growth potential of Dutch fetuses seems to be higher than the potential of fetuses in the Intergrowth study [4–7, 9]. In the future, customized charts may provide a better solution than locally derived population-based charts [13]. When customized charts are introduced, it is important to carefully select the parameters that are taken into account. The parameters should in themselves not be associated with the risk of fetal mortality and morbidity and, if possible, they should be easily and consistently measurable.

The comparison of the Verburg curves with previous- ly used Nicolaides curves demonstrated that all mean values were higher, possibly as a result of a different method used to date pregnancies. Nicolaides used LMP, whilst in our study all pregnancies were dated based on first-tri- mestep CRL [7]. It is known that gestational age tends to be overestimated when LMP is used, resulting in smaller expected values at any given gestational age [14, 15]. Compared with the mean in the studies by Chitty et al. [4–6], the mean HC in our population was small. This discrepancy may reflect that cases were excluded from the Chitty study if at 18–22 weeks menstrual age and ultrasound age differed by more than 10 days.

The Intergrowth charts perform similar to the Ver- burg charts for HC and BPD, but AC and FL of Dutch fetuses seem relatively large. As a result, if we would apply the Intergrowth curves, the percentage of fetuses that would be classified SGA would be relatively small. For the Intergrowth study, Papageorghiou et al. [9] only enrolled pregnancies with a low a priori risk of adverse maternal and perinatal outcome. Moreover, pregnancies with fetal malformation or complications were excluded. The finding that the Dutch fetuses are relatively large may indicate one of two things: (1) Dutch fetuses are destined to be larger than fetuses from populations included in the Intergrowth study, and therefore the Intergrowth charts are not optimal for our region, or (2) Dutch fetuses are relatively healthy and it is acceptable that only a small proportion of fetuses is classified SGA. These findings also apply to the populations studied in Greece and France (Table 3). Whilst it cannot be excluded that in Europe the prevalence of growth-restricted fetuses is low due to relatively high prosperity, it is also possible that the observed differences are related to differences in maternal characteristics [16, 17]. It is of interest that the variations in fetal biometry more or less correspond with variations in maternal size. Findings presented in Table 2 indicate that maternal height in particular is associated with mean birth weight. Further assessment of possible benefit of correcting for maternal size seems indicated.

It is noteworthy that for all fetal measurements the dis- tribution was narrower than expected. The relatively small variation may reflect the improved training and monitor- ing in the Netherlands that is in place since 2007. Stringent criteria have been implemented to ensure that fetal assessments are done with adequate equipment by well-trained personnel. Moreover, before sonographers can participate in the national screening program, they submit a logbook and take a practical exam. Alternatively, the smaller variation may reflect that experienced sonographers have be- come too aware of the “ideal” measurement. A factor that may contribute to the latter is the display of the gestation- al age that corresponds with the obtained measurement on the ultrasound monitor. To examine this hypothesis, a prospective study has been started in which sonographers are blinded to the measurement obtained and the corre- sponding gestational age.

Further research on morbidity and mortality at differ- ent percentile cut-offs is needed to determine whether ei- ther the SD or the cut-off for referral in the case of SGA should be adjusted. In addition, the search for parameters that help distinguish between SGA and growth restriction remains important.

In summary, important steps towards worldwide uni- formity in the assessment of fetal size have been made [9]. Further studies of the importance of fetal and maternal characteristics are warranted. Hereby, priority should be given to parameters that are not associated with morbid- ity and that can be easily measured and unambiguously classified. Whilst we are collecting the data for the above- mentioned studies, the Verburg charts remain recom- mended for the Dutch population.

Acknowledgements

The authors dedicate this paper to the memory of Bero Verburg who passed away at the age of 42 in 2015, many years too soon. We thank sonographers from Verloskundig Echocentrum Alkmaar, Flevoziekenhuis Almere, AMC, BovenIJ Ziekenhuis, DC Kliniek- en OudWest, Noordwest Ziekenhuisgroep Den Helder, OLVG, Verloskunde Centrum Oost Echopunt, Verloskundigen Vida, Rode Kruis Ziekenhuis, Tergooi Ziekenhuis, Verloskundig Echocentrum Midden Kennemerland, Verloskundig Centrum ’t Gooi, Waterland Ziekenhuis, Echocentrum Mama Velserbroek, Echopraktijk WAZ, and Zaans Medisch Centrum for supplying the ultrasound measurements and outcome data. Furthermore, we would like to thank the team of the Dutch Perinatal Registry (Perined) for their assistance in completing outcome details.

Disclosure Statement

The authors report no conflict of interest.
References


