

The impact of choice of reference charts and equations on the assessment of fetal biometry

L. J. SALOMON*, J. P. BERNARD*, M. DUyme†, I. BUvat‡ and Y. VILLE*

*Service de Gynécologie-Obstétrique, Centre Hospitalier Intercommunal de Poissy-St Germain, Poissy, †Laboratoire Biostatistique, Recherche clinique et Santé Publique, Faculté de Médecine de Montpellier, Université Montpellier I, Montpellier and ‡Unité 494, Institut National de la Santé et de la Recherche Médicale, Centre Hospitalier Universitaire Pitié-Salpêtrière, Paris, France

KEYWORDS: charts; fetal biometry; growth; quality; ultrasound; Z-score

ABSTRACT

Objectives The assessment of fetal biometry is usually based on the comparison of measured values with predicted values derived from reference charts or equations in a normal population. This study was undertaken to assess the impact of the choice of reference charts and to develop a Z-score-based tool that could help sonographers to choose the reference charts that best fit their practice.

Methods Fetal biparietal diameter, head circumference, abdominal circumference and femur diaphysis length measurements were made at 20–24 and 30–34 weeks' gestation by four experienced sonographers. All measurements were transformed into Z-scores calculated according to three prediction equations (Snijders and Nicolaidis, 1994; Chitty et al., 1994 and Kurmanavicius et al., 1999). Distributions of Z-scores were compared to the expected standard normal distribution based on mean, SD and Kolmogorov–Smirnov test. Simulations were made to assess sensitivity (Se), specificity (Sp) and Youden's index ($Se + Sp - 1$) of each reference equation, reflecting their ability to identify fetuses with abnormal biometry in our population. The reference that best fitted our practice was determined based on these results.

Results The Z-scores of all biometric parameters were significantly different ($P < 0.001$) when using any of the three reference equations, and none of the Z-score distributions could be considered similar to the standard normal distribution. The number of measurements that would be considered as abnormal according to these references ranged from 2.6% to 23.6%. Se and Sp ranged from 39.59% to 67.12% and 90.14% to 99.69%, respectively.

Conclusion Assessment of fetal biometry is largely dependent on the choice of reference charts. We suggest that the choice of reference charts for fetal biometry could be controlled using Z-scores in each institution and that this could be the first step towards any quality assessment policy. The method we describe for the choice of the most appropriate fetal biometry reference chart might be used for all size charts. Copyright © 2005 ISUOG. Published by John Wiley & Sons, Ltd.

INTRODUCTION

Fetal biometry is an important part of routine examination in the second and third trimesters of pregnancy. Fetal measurements can be combined in order to estimate fetal weight¹ or can be compared to previous measurements in the same fetus in order to evaluate fetal growth longitudinally^{2,3}. However, most measurements are plotted on reference charts for gestation in order to compare fetal measurements with the normal distribution of the reference population. Measurements are considered either adequate, small (i.e. < 3rd, 5th or 10th centile) or large (i.e. > 90th, 95th or 97th centile) and fetal biometry is therefore used as a screening test to identify fetuses that are below or above cut-off values for normality and thus are at increased risk for biometric or morphological abnormalities^{4–6}. Both customization of fetal size charts^{7,8} and the assessment of fetal growth velocity^{3,9} have been developed to improve the ability of fetal biometry to detect high-risk fetuses. However, at the screening level the use of cross-sectional reference charts and equations with the closest distribution to that of the screened population remains the gold standard. This study was undertaken to assess the impact of the choice of reference charts and to develop a Z-score-based tool

Correspondence to: Prof. Y. Ville, Centre Hospitalier Intercommunal de Poissy-St Germain, 10 rue du Champ Gaillard, 78300 Poissy, France (e-mail: yville@wanadoo.fr)

Accepted: 8 February 2005

that can help sonographers to choose the reference charts that best fit their practice.

METHODS

This study was conducted in a population of pregnancies undergoing routine, second- or third-trimester ultrasound examination at 20–24 and 30–34 weeks' gestation, respectively, as part of routine antenatal care in France between June 2001 and January 2004. All measurements were performed to the nearest millimeter with no time constraints by only four trained sonographers using the same probe and ultrasound machine (3.5–5-MHz curvilinear abdominal transducer, General Electric Voluson 730 Expert, GE Medical System Europe-78, Buc, France) with a cine-loop facility. Gestational age (GA), biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and femur diaphysis length (FL) measurements were recorded in all cases. Exclusion criteria were: known abnormal karyotype or congenital malformation, multiple pregnancy or absence of first-trimester dating based on crown–rump length (CRL)¹⁰. No exclusion was made on the basis of abnormal biometry or birth weight.

All biometric measurements were performed according to the methodology published with the reference charts^{11–16}. BPD and HC were measured on a transverse view of the fetal head in an axial plane at the level where the continuous midline echo is broken by the septum pellucidum in the anterior third as described by Campbell and Thoms⁴. BPD was measured with the calipers placed outer to outer. HC was derived from the measurements of the occipital–frontal diameter and the BPD using the formula $\pi(d_1 + d_2)/2$. AC was measured on a transverse circular plane of the fetal abdomen, just above the level of the cord insertion as described by Campbell and Wilkin¹⁷ and was also derived from the two maximum diameters of the circumference. FL was measured on a plane showing the entire femoral diaphysis, with both ends clearly visible and a <45° angle with the horizontal line. At 30–34 weeks, particular care was taken not to include the epiphysis.

All fetal measurements were retrospectively transformed into *Z*-scores according to three different previously published GA-related size charts and equations: References A¹¹, B^{12–14,18} and C^{15,16}. The equations are shown in Appendix S1. All these references should fit our practice based on demographic variables and measurement methodology described in these papers.

In all cases, *Z*-scores were calculated using the formula: $Z\text{-score} = (X_{GA} - M_{GA})/SD_{GA}$, where X_{GA} is the measured value at a known CRL-based GA, M_{GA} is the mean value according to the reference equation used at this GA and SD_{GA} is the SD associated with the mean value at this GA according to the reference equation. Normality of *Z*-score distributions was assessed using the Kolmogorov–Smirnov and Shapiro–Wilk *W*-tests. Due to the large sample sizes, a statistically significant non-normality was accepted unless the normal plot showed a clear deviation from a straight line¹⁸.

Z-score distribution of measurements should follow a standardized normal distribution when fetal measurements performed in our population perfectly fit those of the reference population. Therefore, *Z*-score distributions at 20–24 and 30–34 weeks were analyzed for each of the three reference equations (Reference A, B and C). In all cases, mean and SD values of the *Z*-score distribution were computed. Means of *Z*-scores obtained with each reference were compared to each other using a non-parametric Friedman ANOVA. Means of *Z*-score distributions were tested against the theoretical expected value of 0 using a *t*-test for single sample, SDs were tested against the theoretical expected value of 1 based on the Chi-squared distribution, and the difference between the sample cumulative distribution and the hypothesized standard normal cumulative distribution was assessed using a continuous Kolmogorov–Smirnov one-sample test (Kolmogorov–Smirnov *d* value)^{19,20}.

The measured 5th and 95th centile of the *Z*-score distributions for a given measurement in our population should also fit with the expected 5th and 95th centile values (i.e. –1.645 and +1.645, respectively). Therefore, the actual 5th and 95th centile of the *Z*-score distributions were computed for each type of measurement in order to identify fetuses that were in the tails of our distribution. Simulations were made to compare the number of measurements that would be classified as abnormal as compared to the reference (i.e. *Z*-score < –1.645 or > 1.645).

The sensitivity (Se) and specificity (Sp) of biometry to detect the actual 5th and 95th centile were calculated based on contingency tables (Table S1). The overall number of misclassified fetuses (false-positive + false-negative) was computed for all measurements and all references, and the overall Se, Sp, Youden's index (Se + Sp – 1), and positive (PPV) and negative predictive value (NPV) were calculated at 20–24 and 30–34 weeks both separately and then globally.

The reference (Reference A, B or C) for each type of measurement that best fitted our practice was determined as having a *Z*-score distribution with a mean value and SD closest to 0 and 1, respectively. In addition, it should have the smallest Kolmogorov–Smirnov *d* value, the lowest number of misclassified fetuses and the highest Youden's index for both second- and third-trimester examination taken globally. Indeed, it would be illogical to switch references between the second and third trimesters.

The two-by-two differences between references (References AB, AC and BC) for the prediction of the 5th, 50th and 95th centile were computed and plotted in order to illustrate the difference between all three references across gestation. This was done for BPD as an example.

For all the tests used, a value of $P < 0.05$ was considered statistically significant.

RESULTS

Fetal BPD, HC, AC and FL were measured in 5241 and 4379 ultrasound examinations at 20–24 and 30–34 weeks, respectively, and were included in the

analysis. Z-scores were normally distributed as assessed by normality tests and/or normal plot.

At 20–24 weeks' gestation, the mean and SD of the Z-scores ranged from -1.037 to 0.647 and from 0.746 to 0.926 , respectively. At 30–34 weeks, these ranged from -0.538 to 0.730 and from 0.762 to 1.001 , respectively. Means of Z-scores obtained with each of the three references were statistically different for each type of measurement as assessed by Friedman ANOVA ($P < 10^{-5}$ in all cases). All mean values were statistically different from 0 as assessed by Student's *t*-test ($P \leq 10^{-4}$ in all cases), all SDs were statistically different from 1 based on the Chi-squared distribution ($P \leq 10^{-3}$ in all cases) except Z-score distributions of BPD and HC in the third trimester when assessed upon Reference B¹² ($P = 0.01$ and $P = 0.57$, respectively). Kolmogorov–Smirnov *d* values obtained from the comparison between the actual Z-score distribution and the expected standard normal distribution were significant in all cases ($P < 10^{-2}$) and ranged from 0.062 to 0.432 and from 0.085 to 0.299 in the second and third trimesters, respectively (Table 1).

At 20–24 weeks, measured 5th and 95th centiles of Z-score distributions in our population ranged from -2.438 to -0.607 and 0.329 to 2.145 , respectively. The overall number of fetuses considered to have an abnormal measurement as compared to the expected 5th and 95th centile (-1.645 or 1.645) according to the three reference equations ranged from 174 (3.3%) to 1239 (23.6%).

Se, Sp, PPV and NPV therefore ranged from 32.95% to 64.97%, 79.40% to 100%, 21.63 to 100% and 92.88% to 96.2%, respectively. An illustration of the discrepancy between expected (standardized normal distribution) and measured distribution, generating false-positive screened fetuses, is shown in Figure 1 using an example of a BPD $<$ 5th centile at 20–24 weeks as assessed with Reference C. The number of misclassified fetuses ranged from 186 (3.5%) to 1231 (23.5%) (Table 2).

At 30–34 weeks, measured 5th and 95th centiles of Z-score distributions in our population ranged from -1.997 to -0.771 and 0.8112 to 2.2843 , respectively. The overall number of fetuses considered to have an abnormal measurement as compared to the expected 5th and 95th centile (-1.645 or 1.645) according to the three reference equations ranged from 113 (2.6%) to 717 (16.4%). Se, Sp, PPV and NPV therefore ranged from 25.92% to 79.73%, 87.77% to 100%, 32.78 to 100% and 92.43% to 97.74%, respectively. The number of misclassified fetuses ranged from 187 (4.3%) to 685 (15.6%) (Table 2). Detailed results for screening of fetuses below the 5th centile and above the 95th centile at 20–24 and 30–34 weeks are given in Tables S2 and S3.

Considering both examinations, the number of misclassified fetuses at 20–24 and 30–34 weeks ranged from 424 (4.4%) to 1688 (17.6%). Se, Sp, PPV and NPV ranged from 39.59% to 67.12%, 90.14% to 99.69%, 29.54 to 94.5% and 93.61% to 96.26%, respectively.

Table 1 Mean values and SD of Z-score distributions were tested against 0 and 1, respectively, and the overall distribution was tested against the expected standard normal distribution. Results are shown for the second ($n = 5241$) and third trimesters ($n = 4379$)

Parameter	Reference	Mean	t(t-test)	SD	K–S d value
Second trimester					
BPD	A	-0.090	-7.270^*	0.892^{**}	0.062^{**}
	B	-0.559	-43.716^*	0.926^{**}	0.234^{**}
	C	-1.037	-88.776^*	0.846^{**}	0.432^{**}
HC	A	0.385	35.531^*	0.784^{**}	0.203^{**}
	B	-0.218	-19.531^*	0.808^{**}	0.130^{**}
	C	-0.440	-42.718^*	0.746^{**}	0.237^{**}
AC	A	0.573	54.835^*	0.756^{**}	0.275^{**}
	B	0.579	45.822^*	0.915^{**}	0.232^{**}
	C	0.402	36.008^*	0.809^{**}	0.193^{**}
FL	A	0.647	57.527^*	0.815^{**}	0.297^{**}
	B	0.161	14.676^*	0.794^{**}	0.118^{**}
	C	0.294	26.716^*	0.796^{**}	0.166^{**}
Third trimester					
BPD	A	-0.447	-8.7974^{**}	0.762^{**}	0.229^*
	B	-0.194	-3.1313^{**}	0.978	0.086^*
	C	-0.538	-9.9325^{**}	0.891^{**}	0.239^*
HC	A	-0.054	-0.4229^*	0.801^{**}	0.073^*
	B	0.166	10.9964^{**}	1.002	0.065^*
	C	0.351	24.5354^{**}	0.947^{**}	0.141^*
AC	A	-0.043	-0.8938^*	0.732^{**}	0.089^*
	B	0.730	52.3421^{**}	0.923^{**}	0.299^*
	C	0.677	48.8845^{**}	0.917^{**}	0.280^*
FL	A	0.251	18.0148^{**}	0.923^{**}	0.118^*
	B	0.270	19.4250^{**}	0.919^{**}	0.119^*
	C	0.281	23.3554^{**}	0.795^{**}	0.156^*

* $P < 10^{-2}$. ** $P < 10^{-3}$. Reference A: Snijders and Nicolaides¹¹. Reference B: Chitty *et al.*^{12–14}. Reference C: Kurmanavicius *et al.*^{15,16}. AC, abdominal circumference; BPD, biparietal diameter; FL, femur diaphysis length; HC, head circumference; K–S, Kolmogorov–Smirnov.

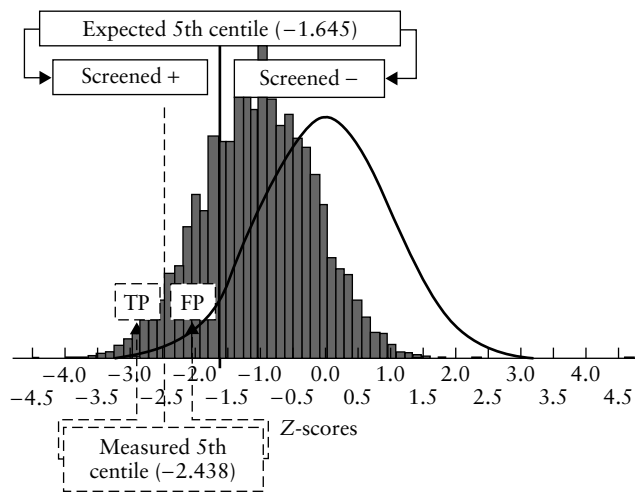


Figure 1 Illustration of discrepancy between expected (standard normal distribution) and measured distribution, generating false-positive screened fetuses. This example is for biparietal diameter below the 5th centile at 20–24 weeks' gestation as assessed with Reference C and could be generalized to all other cases. FP, false-positive; TP, true-positive.

The Youden's index ranged from 0.396 to 0.615. Kolmogorov–Smirnov d values were between 0.045 and

0.335 (Table 3). Based on these results, it appeared that Reference A should be preferred when analyzing BPD measurements and Reference B should be used for HC and FL measurements. For AC measurements, although it appeared that Reference B should not be used, it was not possible to choose between References A and C.

The differences between each couple of references (References AB, AC and BC) for the prediction of 5th, 50th and 95th centile for BPD were plotted and are shown in Figure 2 as an illustration.

DISCUSSION

The use of cross-sectional charts remains the first-line screening tool for growth abnormalities. It is well accepted that the chart used for fetal biometry should be adapted to the population studied. However, to our knowledge this is the first study that extensively examines the impact of the choice of reference charts.

This study did not aim to assess the actual ability of these charts to detect truly abnormally grown fetuses^{21,22}; instead it compared the performance of each reference chart in our practice when measurements were made according to the authors' recommendations. This study

Table 2 The results of screening for both measurements below the 5th and above the 95th centile in the second ($n = 5241$) and third trimesters ($n = 4379$), respectively

Parameter	Reference	Number beyond measured 5th or 95th centile*	Number beyond expected 5th or 95th centile	TP	FP	FN	Number misclassified	Se (%)	Sp (%)	Youden's index	PPV (%)	NPV (%)
Second trimester												
BPD	A	531	345	345	0	186	186	64.97	100	0.650	100	96.20
	B	525	687	297	390	228	618	56.57	91.73	0.483	43.23	94.99
	C	528	1239	268	971	260	1231	50.76	79.40	0.302	21.63	93.50
HC	A	527	325	298	27	229	256	56.55	99.43	0.560	91.69	95.34
	B	527	251	251	0	276	276	47.63	100	0.476	100	94.47
	C	524	280	280	0	244	244	53.44	100	0.534	100	95.08
AC	A	529	419	269	150	260	410	50.85	96.82	0.477	64.20	94.61
	B	527	653	282	371	245	616	53.51	92.13	0.456	43.19	94.66
	C	528	360	278	82	250	332	52.65	98.26	0.509	77.22	94.88
FL	A	520	582	268	314	252	566	51.54	93.35	0.449	46.05	94.59
	B	538	225	225	0	313	313	41.82	100	0.418	100	93.76
	C	528	174	174	0	361	361	32.95	100	0.330	100	92.88
Third trimester												
BPD	A	439	273	237	36	202	238	53.99	99.09	0.531	86.81	95.08
	B	439	448	350	98	89	187	79.73	97.51	0.772	78.13	97.74
	C	438	525	253	272	185	457	57.76	93.10	0.509	48.19	95.20
HC	A	438	166	166	0	272	272	37.90	100	0.379	100	93.54
	B	434	443	343	100	91	191	79.03	97.47	0.765	77.43	97.69
	C	435	440	273	167	162	329	62.76	95.77	0.585	62.05	95.89
AC	A	436	113	113	0	323	323	25.92	100	0.259	100	92.43
	B	438	717	235	482	203	685	53.65	87.77	0.414	32.78	94.46
	C	436	664	238	426	198	624	54.59	89.20	0.438	35.84	94.67
FL	A	430	376	289	87	141	228	67.21	97.80	0.650	76.86	96.48
	B	439	385	280	105	159	264	63.78	97.34	0.611	72.73	96.02
	C	440	360	262	98	220	318	59.55	97.51	0.571	72.78	94.53

*Number of subjects beyond thresholds may vary due to equal results among subjects. Reference A: Snijders and Nicolaides¹¹. Reference B: Chitty et al.^{12–14}. Reference C: Kurmanavicius et al.^{15,16}. AC, abdominal circumference; BPD, biparietal diameter; FL, femur diaphysis length; FN, false-negative; FP, false-positive; HC, head circumference; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity; TP, true-positive.

Table 3 Main overall results for the classification of the cases with the three references at 20–24 and 30–34 gestational weeks taken globally ($n = 9620$)

Parameter	Reference	Mean*	SD*	Number misclassified	Se (%)	Sp (%)	PPV (%)	NPV (%)	Youden's index	K-S d value
BPD	A	-0.25	0.85	424	60.00	99.58	94.17	95.69	0.596	0.132
	B	-0.39	0.97	805	67.12	94.36	57	96.26	0.615	0.164
	C	-0.81	0.90	1688	53.93	85.64	29.54	94.34	0.396	0.335
HC	A	0.19	0.82	528	48.08	99.69	94.5	94.51	0.478	0.108
	B	-0.04	0.92	467	61.81	98.85	85.59	95.89	0.607	0.045
	C	-0.08	0.93	573	57.66	98.07	76.81	95.44	0.557	0.067
AC	A	0.29	0.81	733	39.59	98.27	71.8	93.58	0.379	0.152
	B	0.65	0.92	1301	53.58	90.14	37.74	94.57	0.437	0.261
	C	0.53	0.87	956	53.53	94.13	50.39	94.79	0.477	0.226
FL	A	0.47	0.89	794	58.63	95.37	58.14	95.46	0.54	0.210
	B	0.21	0.86	577	51.69	98.79	82.79	94.76	0.505	0.112
	C	0.31	0.88	679	45.04	98.87	81.65	93.61	0.439	0.160

The reference that best fitted the author's practice for each measurement is shown in bold. *Calculated upon distribution of values at second and third trimesters. Reference A: Snijders and Nicolaides¹¹. Reference B: Chitty *et al.*^{12–14}. Reference C: Kurmanavicius *et al.*^{15,16}. AC, abdominal circumference; BPD, biparietal diameter; FL, femur diaphysis length; HC, head circumference; K-S, Kolmogorov–Smirnov; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity.

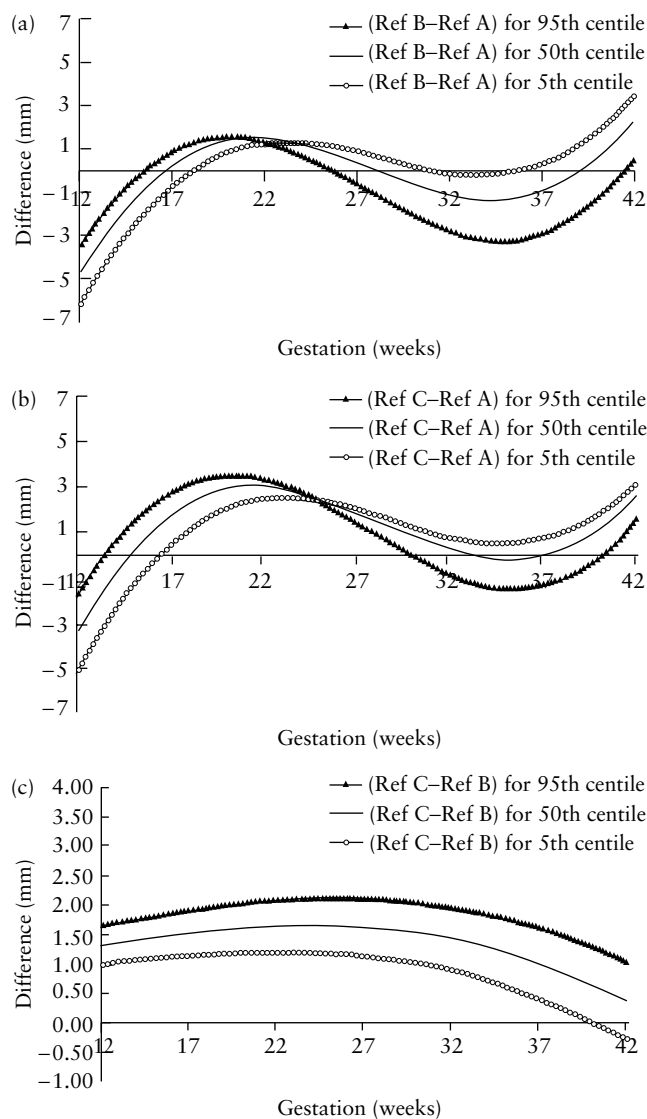


Figure 2 Difference in prediction of the 5th, 50th and 95th centile for biparietal diameter between (a) References A¹¹ and B¹², (b) References A¹¹ and C¹⁵ and (c) References B¹² and C¹⁵.

also did not aim to assess whether one reference was more valid than another. Indeed, all three references were thoroughly derived, using adequate sample size and methodology and taking into account the increasing variability in measurements with gestation^{11–16,18}. Furthermore, although our measurements were performed by trained sonographers who had performed more than 2000 examinations per year for the last 5 years, we did not aim to judge the quality of our database nor to compare sonographers with each other. Indeed, all the sonographers' measurements had narrow normal Z-score distributions with SD values below 1 in almost all cases (Table 1). This is likely to reflect the small number (four) of operators involved in the database and their extensive practice, which is likely to reduce the variability. However, paradoxically, this low variability may actually decrease efficacy as measurements are then compared to inappropriate reference charts. Discrepancies in mean values and in the SD between measurements performed in the studied population and the reference charts are illustrated in Figure 3a and 3b, respectively. Figure 3c and 3d give two examples of acceptable concordance between the actual Z-score distribution and the expected distribution.

Discussion of these results can be divided into three main areas as follows:

- (1) The number of fetuses classified as having a BPD below the expected 5th centile ($Z\text{-score} < -1.645$) at 20–24 weeks ranged from 345 (6.6%) to 1239 (23.7%) when using References A or C, respectively (Table 2). This means that a change in reference chart can lead to a four-fold increase in the risk of being classified as abnormal, leading to mostly unnecessary anxiety and resource allocation for follow-up. A significant proportion of these fetuses may also undergo an invasive procedure and be exposed to a risk of fetal loss. Such false-positive screened fetuses arise from the discrepancy between the measured population and the expected distribution according

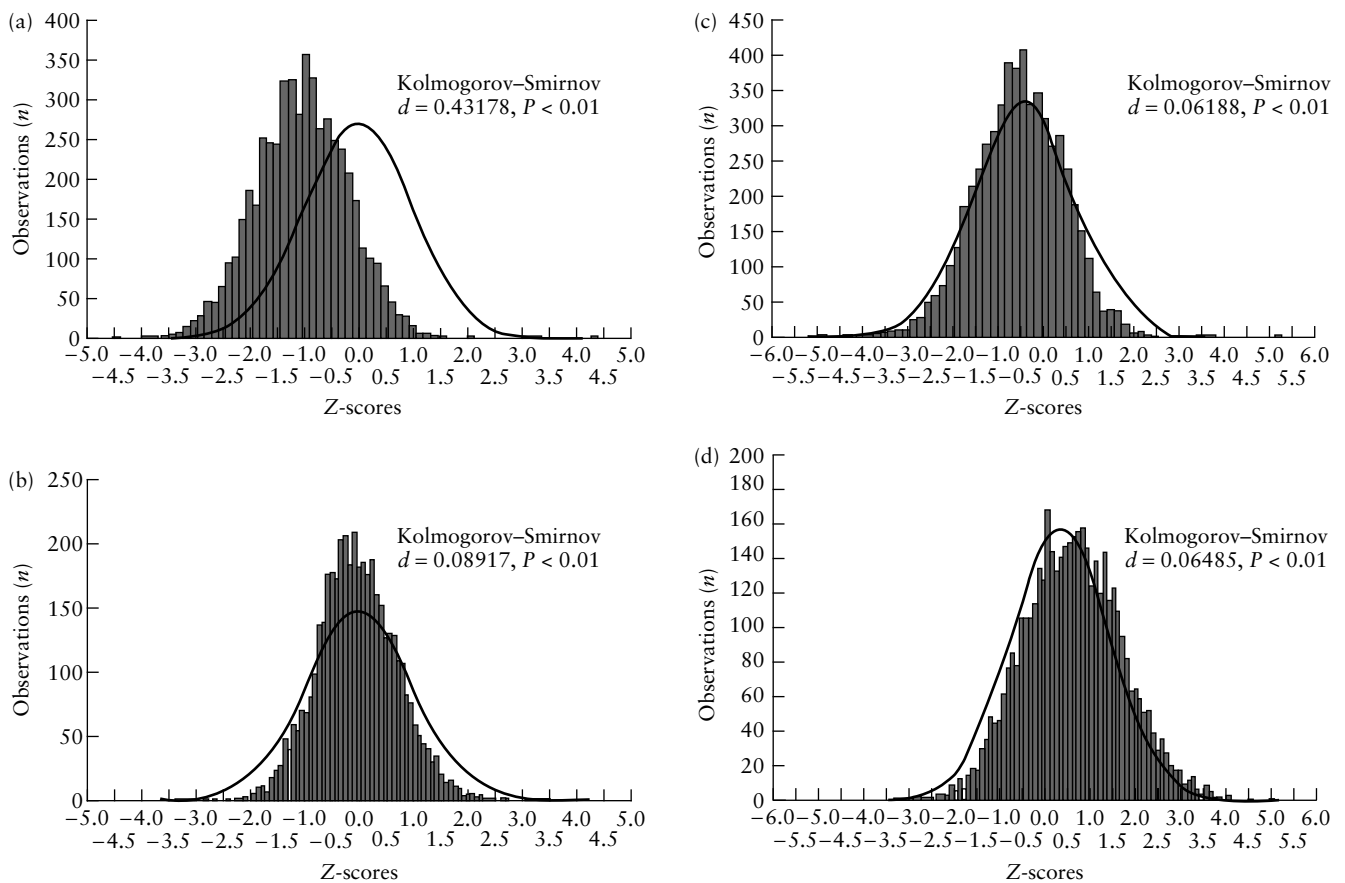


Figure 3 Comparison of Z-score distribution with reference distribution. (a) Distribution of biparietal diameter Z-scores in the second trimester when compared with Reference C¹⁵. Despite the acceptable SD (0.846), the distribution is largely shifted to the left when compared to the expected distribution (mean = -1.037). (b) Distribution of abdominal circumference Z-scores in the third trimester when compared with Reference A¹¹. Although the mean is close to 0 (-0.043) the distribution is too narrow (SD = 0.732), leading to incongruity between our population and the reference. (c) The distribution of biparietal diameter Z-scores in the second trimester when compared with Reference A¹¹ shows an acceptable agreement between our population and the reference (mean = -0.09 , SD = 0.892). (d) The distribution of head circumference Z-scores in the third trimester when compared with Reference B¹² shows an acceptable agreement between our population and the reference (mean = 0.166, SD = 1.002).

to the reference and this is illustrated in Figure 1. Although more fetuses were classified as abnormal with Reference C, this did not lead to an increase in Se or Sp. The same applied when comparing References A and B, leading to a small increase in Se for AC measurement at 30–34 weeks at the expense of a six-fold increase in measurements below the 5th centile. Tables S2 and S3 show that discrepancies affected both the 5th and the 95th centile of our population.

- (2) In our study, Z-score distribution had a mean value and SD that were statistically different in almost all cases from the expected 0 and 1 values, respectively, and Kolmogorov–Smirnov d values were statistically significant in all cases (Table 1). Such unsuitability may lead to inappropriate results in the assessment of fetal size as demonstrated by the poor Se and Sp achieved with some reference charts in our population. This discordance is difficult to control as it may vary throughout gestation, as illustrated by the difference in the results obtained during the second and third trimesters. These variable results throughout gestation are likely to result from differences between

the various references themselves throughout pregnancy (Figure 2). Indeed, References B and C are similar throughout pregnancy but are markedly different from Reference A, and this varies with gestation. However, only one reference should be chosen for each measurement in order to make longitudinal follow-up of fetal growth meaningful.

- (3) This study provides sonographers with useful information on how to choose reference charts and equations that best fit their practice. A preliminary step should consist of transforming all measurements into Z-scores. Z-scores have been increasingly used in recent years and have been designated by the World Health Organization as the recommended system to compare anthropometric measurements to the reference population²³. A major advantage of the Z-score system is that a group of Z-scores can be used as an input for summary statistics such as mean and SD, therefore allowing for the comparison between several groups. If there is good agreement between the observed distribution and the reference distribution then the Z-score distribution should become the standard normal distribution. Although this is only

theoretical, Z-score distribution should have a mean and SD as close as possible to 0 and 1, respectively, and the maximum difference from the expected distribution should be as small as possible as assessed by the Kolmogorov–Smirnov d value. Since biometry is mainly used as a screening test, Se should be as high as possible together with an acceptable Sp in order to avoid unnecessary worry and follow-up. Based on these recommendations and given our measurement distribution, Reference B best fitted our practice for HC reference. BPD and FL measurements should be made using References A and B, respectively. None of these references was found to be acceptable for AC measurements as they all showed poor results in our population.

The assessment of fetal biometry is widely dependent upon the choice of reference charts and equations. Appropriateness of fetal measurements with expected values calculated upon reference equations used in each institution should be controlled and such a process should be the first step towards any quality control policy. Application of Z-scores allows for more accurate use of reference charts and therefore improved identification of at-risk fetuses which in turn should facilitate counseling and make better use of resources.

REFERENCES

- 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**: 333–337.
- Owen P, Donnet ML, Ogston SA, Christie AD, Howie PW, Patel NB. Standards for ultrasound fetal growth velocity. *Br J Obstet Gynaecol* 1996; **103**: 60–69.
- Owen P, Harrold AJ, Farrell T. Fetal size and growth velocity in the prediction of intrapartum Caesarean section for fetal distress. *Br J Obstet Gynaecol* 1997; **104**: 445–449.
- Campbell S, Thoms A. Ultrasound measurement of the fetal head to abdomen circumference ratio in the assessment of growth retardation. *Br J Obstet Gynaecol* 1977; **84**: 165–174.
- Chervenak FA, Jeanty P, Cantraine F, Chitkara U, Venus I, Berkowitz RL, Hobbins JC. The diagnosis of fetal microcephaly. *Am J Obstet Gynecol* 1984; **149**: 512–517.
- Kurtz AB, Goldberg BB. *Obstetrical Measurements in Ultrasound: A Reference Manual*. Year Book Medical Publishers: Chicago, IL, 1988.
- Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992; **339**: 283–287.
- Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995; **6**: 168–174.
- Owen P, Khan KS. Fetal growth velocity in the prediction of intrauterine growth retardation in a low risk population. *Br J Obstet Gynaecol* 1998; **105**: 536–540.
- Wisser J, Dirschedl P, Krone S. Estimation of gestational age by transvaginal sonographic measurement of greatest embryonic length in dated human embryos. *Ultrasound Obstet Gynecol* 1994; **4**: 457–462.
- Snijders RJM, Nicolaides KH. Fetal biometry at 14–40 weeks' gestation. *Ultrasound Obstet Gynecol* 1994; **4**: 34–48.
- Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994; **101**: 35–43.
- Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994; **101**: 125–131.
- Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 4. Femur length. *Br J Obstet Gynaecol* 1994; **101**: 132–135.
- Kurmanavicius J, Wright EM, Royston P, Wisser J, Huch R, Huch A, Zimmermann R. Fetal ultrasound biometry: 1. Head reference values. *Br J Obstet Gynaecol* 1999; **106**: 126–135.
- Kurmanavicius J, Wright EM, Royston P, Zimmermann R, Huch R, Huch A, Wisser J. Fetal ultrasound biometry: 2. Abdomen and femur length reference values. *Br J Obstet Gynaecol* 1999; **106**: 136–143.
- Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br J Obstet Gynaecol* 1975; **82**: 689–697.
- Altman DG, Chitty LS. Charts of fetal size: 1. Methodology. *Br J Obstet Gynaecol* 1994; **101**: 29–34.
- Rosenthal R. An application of the Kolmogorov–Smirnov test for normality with estimated mean and variance. *Psychol Rep* 1968; **22**: 570.
- Ong LD, LeClare PC. The Kolmogorov–Smirnov test for the log-normality of sample cumulative frequency distributions. *Health Phys* 1968; **14**: 376.
- Minior VK, Divon MY. Fetal growth restriction at term: myth or reality? *Obstet Gynecol* 1998; **92**: 57–60.
- Mulik V, Usha Kiran TS, Bethal J, Bhal PS. The outcome of macrosomic fetuses in a low risk primigravid population. *Int J Gynaecol Obstet* 2003; **80**: 15–22.
- World Health Organization (WHO). *Physical Status: The Use and Interpretation of Anthropometry (WHO Technical Report Series No. 854)*. World Health Organization: Geneva, Switzerland, 1995.

SUPPLEMENTARY MATERIAL ON THE INTERNET

The following material is available from the Journal homepage: <http://www.interscience.wiley.com/jpages/0960-7692/suppmat/index.html> (restricted access)

Appendix S1 Prediction equations upon which Z-scores were computed.

Table S1 Method used for the calculation of sensitivity and specificity of biometry to detect fetuses that were actually (a) below the measured 5th centile and (b) above the measured 95th centile.

Table S2 Classification of cases according to the measured distribution and their comparison with the references at 20–24 gestational weeks ($n = 5241$) considering (a) the 5th and (b) the 95th centiles, respectively.

Table S3 Classification of cases according to the measured distribution and their comparison with the references at 30–34 gestational weeks ($n = 4379$) considering (a) the 5th and (b) the 95th centiles, respectively.