Ultrasound Obstet Gynecol 2010; **35**: 426–433
Published online 9 December 2009 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.7518

Birth-weight prediction by two- and three-dimensional ultrasound imaging

J. R. BENNINI, E. F. MARUSSI, R. BARINI, C. FARO and C. F. A. PERALTA

Department of Obstetrics and Gynecology, Center for Integral Assistance to Women's Health, State University of Campinas Medical School, Campinas, Brazil

KEYWORDS: birth-weight estimation; fetal growth; fetal thigh volume; three-dimensional ultrasound; two-dimensional ultrasound

ABSTRACT

Objectives *To compare the accuracies of birth-weight predicting models derived from two-dimensional (2D) ultrasound parameters and from total fetal thigh volumes measured by three-dimensional (3D) ultrasound imaging; and to compare the performances of these formulae with those of previously published equations.*

Methods *A total of 210 patients were evaluated to create a formula-generating group (*n = *150) and a prospective-validation group (*n = *60). Polynomial regression analysis was performed on the first group to generate one equation based on 2D ultrasound measurements, one based on fetal thigh volume measured by the multiplanar technique (ThiM) and one based on fetal thigh volume obtained by the Virtual Organ Computer-aided AnaLysis (VOCAL™) method (ThiV). Paired-samples* t*-tests with Bonferroni adjustments were used to compare the performances of these equations in the formula-finding and the prospective-validation groups. The same approach was used to compare the accuracies of the new 2D and 3D formulae with those of both original and modified 2D equations from previous publications, as well as the 3D model reported by Chang* et al.

Results *The formulae with the best fit for the prediction of birth weight were: estimated fetal weight (EFW) = $-562.824 + 11.962 \times AC \times FDL + 0.009 \times BPD^2 \times AC^2$ (where AC is abdominal circumference, FDL is femur diaphysis length and BPD is biparietal diameter), EFW = $1033.286 + 12.733 \times ThiM$, and EFW = $1025.383 + 12.775 \times ThiV$. For both the formula-generating and the prospective-validation groups, there were no significant differences between the accuracies of the new 2D and 3D models in the prediction of birth weight. When applied to our population, the performances of the modified and original versions of the previously published 2D equations and the performance of the original 3D formula reported by Chang* et al. *were all significantly worse than our models.*

Conclusions *We believe that the greatest sources of discrepancy in estimation of birth weight are the phenotypic differences among patients used to create each of the formulae mentioned in this study. Our data reinforce the need for customized birth-weight prediction formulae, regardless of whether 2D or 3D measurements are employed.* Copyright © 2009 ISUOG. Published by John Wiley & Sons, Ltd.

INTRODUCTION

Accurate estimation of fetal weight is a major concern in perinatal care, because abnormal intrauterine growth is associated with increased neonatal morbidity and mortality[1–4]. For the past 30 years, the assessment of fetal size and growth has essentially been based on predictive models derived from two-dimensional (2D) ultrasound measurements. Although widely used in routine clinical practice, these formulae provide weight estimates with errors of up to 20% when compared with actual birth weights[5–11].

The advent of three-dimensional (3D) ultrasound imaging has allowed the accurate and reliable calculation of fetal organ volumes[12–15]. Some authors have demonstrated that the prediction of birth weight using fetal limb volumetry is more precise than that obtained using conventional 2D ultrasound parameters[15–20]. These studies have compared the accuracy of their new 3D

Correspondence to: Dr C. F. A. Peralta, Departamento de Ginecologia e Obstetrícia, Centro de Atenção Integral à Saúde da Mulher (CAISM), Universidade Estadual de Campinas (UNICAMP), Rua Alexander Fleming, 101 – Cidade Universitária Zeferino Vaz, Distrito de Barão Geraldo, Campinas, S.P., Brasil, CEP 13083-970 (e-mail: cfaperalta@hotmail.com)

Accepted: 7 September 2009

volume-derived equations with that of the traditional 2D measurement models when both are applied to their 3D formula-generating and validation groups. An issue that remains unclear is whether these findings reflect a true improvement offered by 3D ultrasound imaging or simply result from phenotypic differences between the patients used to create each formula.

The aim of this study was to compare the accuracies of 2D and 3D birth-weight predicting models generated from the same sample of patients. In addition, we compared the performances of these formulae with those of previously published 2D and 3D equations when applied to our patients.

METHODS

This was a two-stage prospective cross-sectional study carried out at the Center for Integral Assistance to Women's Health over a 19-month period (between July 2007 and January 2009). The ethics committee of the State University of Campinas Medical School approved this protocol. All patients who agreed to participate signed an informed consent form.

The first phase of the study was carried out over the initial 15 months in order to compose a formula-finding group. The second phase was undertaken during the last 4 months to gather patients for validation of our new birth-weight predicting models. Data from the first subset of women were also used in another report aimed specifically at the comparison between the multiplanar and Virtual Organ Computer-aided AnaLysis (VOCAL™, GE Medical Systems, Zipf, Austria) techniques for the assessment of fetal thigh volume[21].

The eligibility criteria for the entire study were: patients admitted to the hospital for delivery or because of a high probability of the spontaneous onset of labor within the next few days; no maternal diseases or conditions associated with alterations in fetal growth, such as pre-eclampsia, diabetes or tobacco use; singleton pregnancy; well defined gestational age based on the known date of the last menstrual period and/or the measurement of embryonic/fetal crown–rump length during the first trimester, interpreted based on the reference intervals reported by Robinson and Fleming[22]; normal fetal anatomy during obstetric scans and confirmed by postnatal clinical examination; delivery < 49 complete hours after the 2D measurements; and 3D volume acquisition performed at our institution. Exclusion criteria were: multiple pregnancies; the presence of maternal diseases or conditions associated with alterations in fetal growth; uncertain gestational age; fetal anomaly detected by ultrasound imaging or after birth; delivery > 49 complete hours after the 2D and 3D ultrasound evaluations; and delivery in other hospitals. The amount of amniotic fluid was not used as a selection criterion. The patients were non-consecutive because ultrasound examination depended on the weekly schedules of the two physicians. Nonetheless, the inclusion process respected the chronological sequence in which the women were admitted to the hospital.

Demographic characteristics of the mother, including age, gestational age and parity, were recorded at the time of the scan. The subjects included in this study were mostly of mixed race and came from lower socioeconomic backgrounds. These characteristics are representative of the great majority of patients seeking assistance from the public health system of our country. Data on the gestational age at birth, mode of delivery and clinical characteristics of the newborn (weight and Apgar scores) were collected from the mother's hospital records. At our institution, neonates are categorized as small, normal or large for their gestational age if their weights fall below the 10th, between the 10th and the 90th, or above the 90th percentile of the reference intervals defined by Alexander *et al.*[23], respectively. All neonates were weighed immediately after birth in the delivery room on the same precision electronic Filizola Baby scale (Filizola SA, Weighting and Automation, Campo Grande, MS, Brazil), which has a precision of 5 g and is calibrated every 2 weeks. The date and time of each scan and birth were recorded to allow calculation of the time from scan to birth.

All ultrasound examinations were performed transabdominally, with a Voluson 730 Expert scanner, equipped with a RAB 4-8L probe (GE Medical Systems, Milwaukee, WI, USA) by either one or both of two previously selected physicians (J.R.B. and C.F.A.P., with 3 and 10 years' experience with 3D ultrasound imaging, respectively). Examination of the same patient by both operators was limited to cases used for analysis of intraobserver and interobserver variations, which are discussed in another manuscript[21]. Patients who were not included in the assessment of the repeatability and reproducibility of measurements were examined by one physician only. When a woman was evaluated by both physicians, only the first assessment by the first operator (J.R.B.) was used in this analysis.

Two-dimensional ultrasound measurements

Conventional 2D ultrasound fetal biometry was performed as follows. Head measurements were obtained in the axial view at the level of the cavum septi pellucidi, where both thalami could be seen symmetrically, and the anterior and posterior aspects of the cerebral falx were equidistant to the parietal bones. Biparietal diameter (BPD) was measured from the outer edge of the proximal parietal bone to the inner edge of the distal skull table, in a line perpendicular to the orientation of the cerebral falx. The head circumference (HC) was calculated using the scanner's automatically generated ellipse including the outer margins of the fetal skull. Abdominal circumference (AC) was measured in a transverse circular view of the abdomen at the level of the stomach and the portoumbilical vein complex. Anteroposterior (APD) and laterolateral (LLD) diameters across the center of the abdominal image were obtained. The calculation of AC was obtained from the equation $(\text{APD} + \text{LLD}) \times \pi/2$. Femur diaphysis length

(FDL) was measured in a plane in which the full femoral diaphysis was almost parallel to the transducer surface. The measurement was taken from one end of the diaphysis to the other.

Three-dimensional ultrasound measurements

The acquisition and storage of 3D datasets were performed as follows. Initially, the transducer was held over the longitudinal aspect of the fetal femur. Several thigh volumes were acquired using automatic sweeps, but two to four of them were stored for further analysis. The 3D volume box was adjusted for the size of the thigh, and the sweep angle was set at between 30 and 70° depending on the gestational age. The slowest sweep velocity (four frames per s) was chosen in order to guarantee the best resolution of the image, and the acquisition process was repeated if there was any maternal or fetal movement. Volumes affected by motion artifacts were not stored.

Thigh volumetry using both the multiplanar and VOCAL techniques was performed offline with 4D view software version 5.3 (GE Medical Systems) on datasets retrieved from the scanner by the same operator who examined the patient. For this purpose, the physician selected only one dataset among all those stored for a specific patient, on the basis of image quality, and measured thigh volumes using both methods (multiplanar and VOCAL).

Total fetal thigh volumetry by the multiplanar technique was performed as previously described by Chang *et al.*[16] and by ourselves[21]. Briefly, the surface of the thigh was drawn manually in every axial view 3 mm apart from each other, from one end of the femur diaphysis to the other. At the end of this process, the total thigh volume (ThiM) was automatically calculated by the built-in scanner software.

Thigh volumetry by the VOCAL technique was performed as previously described by our group[21]. The dataset containing the fetal thigh was displayed on the screen in three orthogonal planes. The sagittal view of the femur was exposed in one of these planes, and this image was rotated so that the thigh and whole diaphysis were identified in a perpendicular position. Two demarcating arrows were positioned at each end of the diaphysis to define the limits of the thigh to be included in the volume calculation. Volume estimates (ThiV) were computed utilizing the VOCAL program with a manual trace at 30° of rotation.

Statistical analysis

Comparison of the demographic characteristics of the formula-generating and prospective-validation groups

Patients in the formula-generating group were compared with those from the prospective-validation group with regard to age, gestational age at the time of the scan, parity, mode of delivery, time from scan to birth, weight of the neonate, Apgar scores and neonatal weight category (small-, normal- or large-for-gestational age). Independent-samples *t*-tests and Chi-square tests were used for the assessment of continuous and categorical variables, respectively. The Mann–Whitney *U*-test was used for comparison of Apgar scores between the two groups. Two-tailed $P < 0.05$ was considered statistically significant.

Creation of new two- and three-dimensional ultrasound models

Using data from the first phase of the study and weight of the neonate as the dependent variable, polynomial stepwise regression analyses up to the third order were considered to generate three birth-weight predicting models with the following predictors: conventional 2D ultrasound measurements, total fetal thigh volume measured using the multiplanar method (ThiM) and total fetal thigh volume measured by the VOCAL technique (ThiV). The formulae including thigh volumetry were also used in another report of the comparison of these methods[21]. To construct the 2D formula, eigenvalue, tolerance, variance inflation factor, condition index and variance proportion were calculated to check for multicollinearity among independent variables[24]. The criteria for multicollinearity were: eigenvalue < 0.1, tolerance value $< 1 - r^2$, variance inflation factor $> 1/(1 - r^2)$, condition index > 0.30 and variance proportion > 0.8. These tests were not performed when the 3D measurement models were derived, as in such cases only the first, second and third powers of thigh volumes were evaluated as independent variables. For all best-fit equations, Kolmogorov–Smirnov tests were performed to check for normality of the standardized residuals.

Comparison of previously published equations with the models created in this study

The original and modified 2D ultrasound formulae described by Hadlock *et al.*[8,9], Woo *et al.*[10] and Hsieh *et al.*[11], as well as the original 3D ultrasound equation reported by Chang *et al.*[16], were applied to our data in order to compare their performances with those of our new models. The criterion for the selection of these specific 2D equations was the presence of the same parameters (BPD, AC and FDL) used in our new 2D formula. The reason for choosing Chang's 3D model was that we used the same multiplanar technique for the measurement of total fetal thigh volume. The modified 2D ultrasound models (obtained from the original formulae published by Hadlock *et al.*[8,9], Woo *et al.*[10] and Hsieh *et al.*[11]) were generated by the calculation of new coefficients derived from multiple regression analysis using their equations' terms and our data. By using this approach, we sought to minimize potential biases in comparing the prediction errors of functions generated from two different study populations. The original and modified formulae used for comparison with our new models are described in Table 1.

Table 1 Original and modified two-dimensional (2DUS) and three-dimensional (3DUS) ultrasound imaging birth-weight prediction models used for comparison with the new formulae created in this study

Formula	*Parameter*	*Equation*
Hadlock *et al.*[8]	2DUS	$\text{Log}_{10}\ \text{EFW} = 1.3598 + 0.051 \times \text{AC} + 0.1844 \times \text{FDL} - 0.0037 \times \text{AC} \times \text{BPD}$
Modified Hadlock *et al.*[8]	2DUS	$\text{Log}_{10}\ \text{EFW} = 2.368 + 0.011 \times \text{AC} + 0.068 \times \text{FDL} + 0.001 \times \text{AC} \times \text{BPD}$
Hadlock *et al.*[9]	2DUS	$\text{Log}_{10}\ \text{EFW} = 1.335 - 0.0034 \times \text{AC} \times \text{FDL} + 0.0316 \times \text{BPD} + 0.0457 \times \text{AC} + 0.1623 \times \text{FDL}$
Modified Hadlock *et al.*[9]	2DUS	$\text{Log}_{10}\ \text{EFW} = 1.252 - 0.004 \times \text{AC} \times \text{FDL} + 0.027 \times \text{BPD} + 0.048 \times \text{AC} + 0.192 \times \text{FDL}$
Woo *et al.*[10]	2DUS	$\text{Log}_{10}\ \text{EFW} = 1.54 + 0.15 \times \text{BPD} + 0.00111 \times \text{AC}^2 - 0.0000764 \times \text{BPD} \times \text{AC}^2 + 0.05 \times \text{FDL} - 0.000992 \times \text{FDL} \times \text{AC}$
Modified Woo *et al.*[10]	2DUS	$\text{Log}_{10}\ \text{EFW} = 1.794 + 0.080 \times \text{BPD} + 0.001 \times \text{AC}^2 - 0.000046 \times \text{BPD} \times \text{AC}^2 + 0.126 \times \text{FDL} - 0.002 \times \text{FDL} \times \text{AC}$
Hsieh *et al.*[11]	2DUS	$\text{Log}_{10}\ \text{EFW} = 2.7193 + 1.745 \times 0.001 \times \text{BPD}^2 \times \text{FDL} - 7.6742 \times 0.0001 \times \text{AC} \times \text{BPD}^2 - 0.1432 \times \text{FDL} + 9.4962 \times 0.001 \times \text{AC} \times \text{BPD}$
Modified Hsieh *et al.*[11]	2DUS	$\text{Log}_{10}\ \text{EFW} = 2.369 - 0.00012 \times \text{BPD}^2 \times \text{FDL} - 0.000096 \times \text{AC} \times \text{BPD}^2 + 0.077 \times \text{FDL} + 0.003 \times \text{AC} \times \text{BPD}$
Chang *et al.*[16]	3DUS	$\text{EFW} = 1080.8735 + 22.44701 \times \text{ThiM}$

AC, abdominal circumference (in cm); BPD, biparietal diameter (in cm); EFW, estimated fetal weight (in g); FDL, femur diaphysis length (in cm); ThiM, fetal thigh volume measured by the multiplanar technique (in cm^3).

In both the formula-generating and the prospective-validation groups, the performances of each of these equations were analyzed by the calculation of systematic and random errors. The systematic error, or accuracy, was evaluated by calculating the mean sign percentage error ((estimated fetal weight − actual birth weight)/actual birth weight × 100). The random error, or precision, was evaluated by calculating the SD of the mean sign percentage error. Paired-samples *t*-tests with Bonferroni adjustments were used to detect significant differences between the accuracies of these formulae. The adjusted *P* value (P'), which was calculated according to the Bonferroni method, was obtained from the formula $P' = k \times P$, where k was the number of paired comparisons and the *P* value was obtained from each paired-samples *t*-test[25,26]. In this manner, for the comparison of the accuracies of our new 2D and 3D formulae, the P' for each paired comparison was obtained by the formula $3 \times P$ value. Similarly, for the comparison of our 2D model with the original and modified functions of Hadlock *et al.*[8,9], Woo *et al.*[10] and Hsieh *et al.*[11] (eight paired comparisons), each P' was obtained by the formula $P' = 8 \times P$ value of the paired-samples *t*-test. This method has the restriction that the P' cannot exceed 1.0. In order to compare the random errors of two equations, correlated variance tests for paired samples were used[27]. For each paired comparison, the variances were considered to be significantly different if the *P* value obtained from the *r* (Pearson's correlation coefficient) distribution table was less than 0.05. The *r* value was calculated using the formula: $r = (F - 1)/\sqrt{((F+1)^2 - 4 \times r^2 \times F)}$, where *F* is the ratio of the variances of the groups being compared. All *P* values exceeding 0.200 were referred to as $P > 0.200$.

The data were analyzed using the statistical software packages SPSS 16.0 (SPSS, Chicago, IL, USA) and Excel for Windows 2007 (Microsoft Corp., Redmond, WA, USA).

RESULTS

A total of 254 patients were evaluated throughout the whole study period. The first 182 women were examined during the initial 15 months (Phase 1) with the purpose of composing a formula-generating group. Of these, 153 met the entry criteria. A total of 29 cases were excluded because their deliveries occurred more than 49 h after the scan. During the last 4 months of the study (Phase 2), another 72 patients were evaluated in order to create a prospective-validation group. Ten of these were excluded because their deliveries occurred more than 49 h after the scan.

Of the 153 patients from the formula-generating group, three more were withdrawn from final analysis because fetal thigh volumetry was not possible owing to poor image quality in the 3D dataset. For the same reason, two of the 62 cases in the prospective-validation group were eliminated. Demographic and clinical data for the remaining 150 patients from the first group and 60 patients from the second group are presented in Table 2. No statistically significant differences were noted between these groups with regard to maternal age, gestational age at the time of the scan, parity, mode of delivery, time from scan to birth, weight of the neonate, Apgar scores or neonatal weight category.

The best-fit formulae for the estimation of fetal weight using 2D ultrasound parameters and 3D fetal thigh volumes were: estimated fetal weight (EFW) = $-562.824 + 11.962 \times \text{AC} \times \text{FDL} + 0.009 \times \text{BPD}^2 \times \text{AC}^2$ (SD of predicted values, 554.261; $r = 0.899$; $r^2 = 0.808$; $P < 0.001$), $\text{EFW} = 1033.286 + 12.733 \times \text{ThiM}$ (SD of predicted values, 566.521; $r = 0.919$; $r^2 = 0.845$; $P < 0.001$) and $\text{EFW} = 1025.383 + 12.775 \times \text{ThiV}$ (SD of predicted values, 570.6299; $r = 0.926$; $r^2 = 0.857$; $P < 0.001$). The results of the Kolmogorov–Smirnov tests revealed normal distributions of the standardized residuals for all equations.

For both the formula-generating and the prospective-validation groups, no statistically significant differences

Table 2 Demographic characteristics of the formula-generating and prospective-validation groups

Characteristics	*Formula-generating group (n = 150)*	*Prospective-validation group (n = 60)*	P
Mother			
Age (years)	26.4 ± 6.3 (15–43)	27.1 ± 6.1 (16–43)	0.535*
Gestational age (weeks)	38.4 ± 2.3 (29.7–41.7)	38.5 ± 5.8 (29.4–42.4)	0.970*
Parity			0.457†
Nulliparous	47	22	
Parous	103	38	
Mode of delivery			0.629†
Vaginal	68 (45.3)	25 (41.7)	
Cesarean section	82 (54.7)	35 (58.3)	
Time between scan and birth (h)	18.1 ± 14.7 (0.7–48.6)	19.1 ± 16.0 (0.7–48)	0.676*
Neonate			
Weight (g)	3124.7 ± 616.5 (1445–4500)	3247.4 ± 698.3 (1475–4750)	0.211*
< 2500 g	21 (14)	7 (11.7)	0.567†
2500–4000 g	118 (78.7)	46 (76.7)	
> 4000 g	11 (7.3)	7 (11.7)	
Apgar score			
1 min	9 (2–10)	9 (2–10)	0.983‡
5 min	10 (6–10)	10 (7–10)	0.718‡
Size			0.909†
Small	16 (10.7)	7 (11.7)	
Adequate	119 (79.3)	46 (76.7)	
Large	15 (10.0)	7 (11.7)	

Values are mean ± SD (range), *n* (%) or median (range). *Independent-samples *t*-test. †Chi-square test. ‡Mann–Whitney *U*-test.

Table 3 Accuracy and precision of birth-weight prediction of the two-dimensional (2D) and three-dimensional (3D) equations created in this study in the formula-generating and prospective-validation groups

		Versus 2D formula		*Versus 3D formula (ThiM)*	
Group/Formula	*MSPE ± SD (%)*	P′ *MSPE*	P *SD*	P′ *MSPE*	P *SD*
Formula-generating group					
New 2D formula	−0.27 ± 8.35				
New 3D formula (ThiM)	0.69 ± 7.64	0.594	> 0.200		
New 3D formula (ThiV)	0.73 ± 7.99	0.627	> 0.200	1.000	0.100
Prospective-validation group					
New 2D formula	1.27 ± 8.78				
New 3D formula (ThiM)	−0.82 ± 8.24	0.402	> 0.200		
New 3D formula (ThiV)	0.06 ± 8.08	1.000	> 0.200	0.087	> 0.200

MSPE ± SD, mean sign percentage error (accuracy) ± SD (precision) of MSPE; P′ MSPE, paired-samples *t*-tests with Bonferroni adjustment for comparison of MSPE; *P* SD, correlated variance in paired samples for comparison of SD; ThiM, thigh volume measured using multiplanar technique; ThiV, thigh volume measured using VOCAL technique.

were noted between the accuracies and precisions of our 2D and 3D models in the prediction of birth weight (Table 3).

The performances of previously published formulae, when applied to our population, were significantly worse than those of our new models (Tables 4 and 5). Although our 2D function allowed the estimation of birth weight with a mean ± SD sign percentage error of −0.27 ± 8.35% in the formula-finding group, the result obtained from the original model reported by Hsieh *et al.*[11] was 5.97 ± 15.34%. The performances of all other original and modified 2D models were even worse than the model described by Hsieh *et al.*[11] (Table 4). In the prospective-validation group, our 2D equation had an accuracy of 1.27%, which was not significantly different from those obtained using the original and modified formulae of Hsieh *et al.*[11] and the original models reported by Woo *et al.*[10] and Hadlock *et al.*[9] (Table 4). However, the random errors produced by all formulae tested in this study were significantly worse than those resulting from our new 2D model, in both the formula-finding and the prospective-validation groups. Similarly, the performance of our new 3D function that used ThiM was significantly better than that of Chang's equation, in both the formula-finding and the prospective-validation groups (Table 5).

Table 4 Accuracy and precision of birth-weight prediction of the two-dimensional (2D) equation created in this study and of the original and modified previously published 2D models

		Hsieh et al.[11]		*Woo* et al.[10]		*Hadlock* et al.[8]		*Hadlock* et al.[9]	
Parameter	*New 2D formula*	*Original*	*Modified*	*Original*	*Modified*	*Original*	*Modified*	*Original*	*Modified*
Formula-generating group									
MSPE ± SD (%)	−0.27 ± 8.35	5.97 ± 15.34	7.94 ± 16.07	6.06 ± 15.19	18.72 ± 16.04	−12.43 ± 15.35	11.25 ± 16.18	7.64 ± 15.68	10.13 ± 15.96
P′ MSPE		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
P SD		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Prospective-validation group									
MSPE ± SD (%)	1.27 ± 8.78	5.71 ± 17.72	7.23 ± 17.61	5.73 ± 17.47	18.41 ± 17.51	−13.76 ± 14.81	10.65 ± 17.64	6.71 ± 16.69	8.99 ± 17.25
P′ MSPE		0.336	0.056	0.376	< 0.001	< 0.001	< 0.001	0.064	< 0.001
P SD		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

MSPE ± SD, mean sign percentage error (accuracy) ± SD (precision) of MSPE; *P′* MSPE, paired-samples *t*-tests with Bonferroni adjustment for comparison of MSPE; *P* SD, correlated variance in paired samples for comparison of SD.

Table 5 Comparison of the accuracy and precision of a previously published three-dimensional (3D) formula with those of our new 3D equation using thigh volumetry as determined by the multiplanar method

Formula	*Formula-generating group*	*Prospective-validation group*
New 3D formula (ThiM) (MSPE ± SD) (%)	0.73 ± 8.00	−0.82 ± 8.24
Chang *et al.*[11] formula (MSPE ± SD) (%)	52.75 ± 12.27	50.64 ± 12.40
P MSPE	< 0.001	< 0.001
P SD	< 0.001	< 0.001

MSPE, mean sign percentage error (accuracy) ± SD (precision) of MSPE; *P* MSPE, paired-samples *t*-tests for comparison of MSPE; *P* SD, correlated variance in paired samples for comparison of SD; ThiM, thigh volume measured using multiplanar technique.

DISCUSSION

This study shows that there is no statistically significant difference in the prediction of birth weight by 2D and 3D formulae that are generated from the same set of patients. Conversely, it demonstrates that 2D and 3D equations that were created using different populations have significantly lower performances in the prediction of birth weight in our patients than our models.

The main question raised in this study was whether the superior accuracy of 3D formulae over previously published 2D equations reflects a true advantage of 3D ultrasound imaging, or whether it is simply a result of phenotypic differences between the patients used to create each of these formulae. Chang *et al.*[16] concluded that 3D ultrasound assessment of fetal thigh volume has better accuracy in predicting birth weight than 2D ultrasound formulae previously produced by Warsof *et al.*[5], Thurnau *et al.*[7] and Hadlock *et al.*[9]. The 3D equation yielded a mean value of error of 0.0 g, percentage error of 0.7%, absolute error of 176.1 g and absolute percentage error of 5.8%, whereas the results for the same parameters using the 2D formulae were −206.2 g, −6.2%, 249.0 g and 7.5% (Warsof *et al.*[5]), −708.4 g, −20.8%, 708.4 g and 20.8% (Thurnau *et al.*[7]), and −224.9 g, −6.7%, 260.1 g and 7.8% (Hadlock *et al.*[9]). These values were all significantly different from those obtained with Chang's equation. Similarly, Lee *et al.*[20] demonstrated that a different way of measuring fetal thigh volume, the fractional limb volume, can also be used to predict birth weight with better accuracy than the formulae reported by Shepard *et al.*[6] and Hadlock *et al.*[9] Other authors have used fetal upper arm volumes[17] and abdominal organ volumes[18] to estimate birth weight, and have also demonstrated an improved performance of these formulae compared with those of traditional 2D equations. Our results, at least in part, contradict these findings. We observed that, if 2D and 3D formulae are generated from the same set of patients, they have similar accuracies for the prediction of birth weight. More recently, in 176

prolonged pregnancies, Lindell and Marsal demonstrated that fetal weight could be estimated using 2D sonography with the same accuracy as 3D sonography[28].

In this study, we compared the accuracies of the most commonly used 2D formulae and of the 3D equation reported by Chang *et al.*[16] with those of our new 2D and 3D models. We did not compare our 3D equation with the formulae generated by Song *et al.*[19] and Lee *et al.*[20] because the techniques used to calculate fetal thigh volume were different. Comparisons with the equations created by Schild *et al.*[18] were not possible either, because they used fetal abdominal organ volumes in association with thigh volumes to predict weight. The selection of conventional 2D formulae for comparison with our models was based on the predictor variables included in these functions. We chose only the formulae that used BPD, FDL and AC, which were the parameters used in our equation. In addition, we took care to generate modified 2D ultrasound models from the original formulae published by Hadlock *et al.*[8,9], Woo *et al.*[10] and Hsieh *et al.*[11] for comparison with our 2D equation. As mentioned previously in this article, by using this approach, we sought to minimize potential biases in comparing the prediction errors of formulae generated from two different study populations.

Our results support the claim that birth-weight predicting formulae should be customized to each specific population. We noticed striking differences in the prediction of birth weight between our formulae and others. When we applied the original equations reported by Hsieh *et al.*[11], Woo *et al.*[10] and Hadlock *et al.*[8,9] to the patients in our formula-finding group, we obtained corresponding mean ± SD sign percentage errors of $5.97 \pm 15.34\%$, $6.06 \pm 15.19\%$, $-12.43 \pm 15.35\%$ and $7.64 \pm 15.68\%$. These results were all significantly different from those obtained using our new 2D equation (P' MSPE < 0.001; P SD < 0.001). The performances of the modified models were even worse when applied to the formula-generating group. Regarding the prospective-validation group, there were no differences observed in the accuracies of the original and modified formulae of Hsieh *et al.*[11], the original model of Woo *et al.*[10] or the original model of Hadlock *et al.*[9] and the accuracy of our 2D equation. However, in this group, these formulae yielded significantly higher random errors than those generated by our model. Moreover, it is important to consider that we used paired-samples *t*-tests with Bonferroni adjustment for the comparison of the accuracies of different equations, in both the formula-finding and the prospective-validation groups. This statistic was chosen instead of ANOVA because our intention was to compare specific pairs of observations rather than all possible combinations of groups of measurements. This approach is known to be highly conservative in controlling the overall Type I error when multiple comparisons are made. Therefore, we were extremely rigid in considering all possible original and modified formulae (we used a total of eight different equations) for comparison with our 2D model, as the corrected P values obtained for each paired comparison resulted from the multiplication of each P value derived from the paired samples *t*-test by eight. In their article regarding fractional limb volume, Lee *et al.*[20] also applied a modified formula generated by Hadlock *et al.*[9] to their patients and noticed a better performance than that obtained with the original equation. In our evaluation, the modified models of all previously published 2D equations did worse than the original models at predicting birth weight in our population. When we compared our 3D equation to that of Chang *et al.*[16], the differences were even greater than those observed in the comparison of our 2D equation with the other 2D models.

One may argue that our findings are the result of inaccurately performed 2D and 3D measurements. We believe this hypothesis to be extremely unlikely because the 2D ultrasound measurements of BPD, HC, AC and FDL have been employed universally for years, and the physicians who evaluated the patients in this study have significant experience in performing obstetric ultrasound examinations. The greater discrepancy found in the comparison of the 3D formulae may have resulted from a systematic difference in the method of fetal thigh volumetry because the training process to become experienced with 3D ultrasound imaging and fetal organ volumetry is more complex and time consuming. All of the steps described by Chang *et al.*[16], however, were followed in our patients. Moreover, before data acquisition for this analysis, both sonographers involved in fetal thigh volumetry underwent a training process and each evaluated 50 datasets. In addition, in a separate analysis, we demonstrated that both the repeatability and reproducibility of total fetal thigh volume measurements using the multiplanar technique are acceptable, and are comparable with the results obtained by other authors[20,21].

In conclusion, we believe that the greatest sources of discrepancy in estimation of birth weight are the phenotypic differences among the patients used to create each of the formulae mentioned in this study. We suppose that lower errors would result from the use of simpler processes, which could give 2D measurements an advantage. In addition, we believe that our data reinforce the need for customization of birth-weight prediction formulae, regardless of whether 2D or 3D measurements are employed.

REFERENCES

1. Sung IK, Vohr B, Oh W. Growth and neurodevelopmental outcome of very low birth weight infants with intrauterine growth retardation: comparison with control subjects matched by birth weight and gestational age. *J Pediatr* 1993; **123**: 618–624.
2. Bardin C, Zelkowitz P, Papageorgiou A. Outcome of small-for-gestational age and appropriate-for-gestational age infants born before 27 weeks of gestation. *Pediatrics* 1997; **100**: E4.
3. Roth S, Chang TC, Robson S, Spencer JA, Wyatt JS, Stewart AL. The neurodevelopmental outcome of term infants with different intrauterine growth characteristics. *Early Hum Dev* 1999; **55**: 39–50.

4. Dashe JS, McIntire DD, Lucas MJ, Leveno KJ. Effects of symmetric and asymmetric fetal growth on pregnancy outcomes. *Obstet Gynecol* 2000; **96**: 321–327.
5. Warsof SL, Gohan P, Berkowitz RL, Hobbins JC. The estimation of fetal weight by computer-assisted analysis. *Am J Obstet Gynecol* 1977; **128**: 881–892.
6. Shepard MJ, Richards VA, Berkowitz RL, Warsof SL, Hobbins JC. An evaluation of two equations for predicting fetal weight by ultrasound. *Am J Obstet Gynecol* 1982; **142**: 47–54.
7. Thurnau GR, Tamura RK, Sabbagha R, Depp OR, Dyer A, Larkin R, Lee T, Laughin C. A simple estimated fetal weight equation based on real-time ultrasound measurements of fetuses less than thirty-four weeks' gestation. *Am J Obstet Gynecol* 1983; **145**: 557–561.
8. Hadlock FP, Harrist RB, Carpenter RJ, Deter RL, Park SK. Sonographic estimation of fetal weight. The value of femur length in addition to head and abdomen measurements. *Radiology* 1984; **150**: 535–540.
9. Hadlock FP, Harrist RB, Shearman 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.
10. Woo JS, Wan CW, Cho KM. Computer-assisted evaluation of ultrasonic fetal weight prediction using multiple regression equations with and without the fetal femur length. *J Ultrasound Med* 1985; **4**: 65–67.
11. Hsieh FJ, Chang FM, Huang HC, Lu CC, Ko TM, Chen HY. Computer-assisted analysis for prediction of fetal weight by ultrasound: comparison of biparietal diameter, abdominal circumference and femur length. *J Formosan Med Assoc* 1987; **86**: 957–964.
12. Chang CH, Chang FM, Yu CH, Ko HC, Chen HY. Assessment of fetal cerebellar volume using three-dimensional ultrasound. *Ultrasound Med Biol* 2000; **26**: 981–988.
13. Peralta CF, Cavoretto P, Csapo B, Falcon O, Nicolaides KH. Lung and heart volumes by three-dimensional ultrasound in normal fetuses at 12–32 weeks' gestation. *Ultrasound Obstet Gynecol* 2006; **27**: 128–133.
14. Boito SME, Laudy JAM, Struijk PC, Stijnen T, Wladimiroff JW. Three-dimensional US assessment of hepatic volume, head circumference, and abdominal circumference in healthy and growth-restricted fetuses. *Radiology* 2002; **223**: 661–665.
15. Lee W, Balasubramaniam M, Deter RL, Hassan SS, Gotsch F, Kusanovic JP, Gonçalves LF, Romero R. Fractional limb volume – a soft tissue parameter of fetal body composition: validation, technical considerations and normal ranges during pregnancy. *Ultrasound Obstet Gynecol* 2009; **33**: 427–440.
16. Chang FM, Liang RI, Ko HC, Yao BL, Chang CH, Yu CH. Three-dimensional ultrasound-assessed fetal thigh volumetry in predicting fetal weight. *Obstet Gynecol* 1997; **90**: 331–339.
17. Liang RI, Chang FM, Yao BL, Chang CH, Yu CH, Ko HC. Predicting birth weight by fetal upper-arm volume with use of three-dimensional ultrasonography. *Am J Obstet Gynecol* 1997; **177**: 632–638.
18. Schild RL, Fimmers R, Hansmann M. Fetal weight estimation by three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2000; **16**: 445–452.
19. Song TB, Moore TR, Lee JY, Kim YH, Kim EK. Fetal weight prediction by thigh volume measurement with three-dimensional ultrasonography. *Obstet Gynecol* 2000; **96**: 157–161.
20. Lee W, Deter RL, Ebersole JD, Huang R, Blanckaert K, Romero R. Birth weight prediction by three-dimensional ultrasonography – fractional limb volume. *J Ultrasound Med* 2001; **20**: 1283–1292.
21. Bennini JR, Faro C, Marussi EF, Barini R, Peralta CFA. Fetal thigh volumetry by three-dimensional ultrasound: comparison between multiplanar and VOCAL™ techniques. *Ultrasound Obstet Gynecol* 2010. DOI: 10.1002/uog.7517.
22. Robinson HP, Fleming JE. A critical evaluation of sonar 'crown–rump length' measurements. *Br J Obstet Gynaecol* 1975; **82**: 702–710.
23. Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. *Obstet Gynecol* 1996; **87**: 163–168.
24. Liu RX, Kuang J, Gong Q, Hou XL. Principal component regression analysis with SPSS. *Comput Methods Programs Biomed* 2003; **71**: 141–147.
25. Brown BW, Russel K. Methods of correcting for multiple testing: operating characteristics. *Stat Med* 1997; **16**: 2511–2528.
26. Ludbrook J. Multiple comparison procedures update. *Clin Exp Pharmacol Physiol* 1998; **25**: 1032–1037.
27. Pitman EJG. A note on normal correlation. *Biometrika* 1939; **31**: 9–12.
28. Lindell G, Marsal K. Sonographic fetal weight estimation in prolonged pregnancy: comparative study of two- and three-dimensional methods. *Ultrasound Obstet Gynecol* 2009; **33**: 295–300.

