Influence of ultrasound-to-delivery interval and maternal–fetal characteristics on validity of estimated fetal weight

J. M. COHEN*, J. A. HUTCHEON†, M. S. KRAMER†‡, K. S. JOSEPH§, H. ABENHAIM¶
and R. W. PLATT*†

Department of Biostatistics, The Montreal Children's Hospital Research Institute, Departments of †Epidemiology, Biostatistics and Occupational Health and ‡Pediatrics, McGill University and ¶Department of Obstetrics and Gynecology, SMBD Jewish General Hospital, Montreal, QC and §Department of Obstetrics and Gynecology and Pediatrics, Dalhousie University and the IWK Health Centre, Halifax, Nova Scotia, Canada

ABSTRACT

Objectives *To explore the effects of ultrasound-to-delivery interval and maternal–fetal characteristics on the distribution of measurement error in estimated fetal weights (EFWs), and to determine the predictive ability of EFW for diagnosis of small-for-gestational age (SGA) and large-for-gestational age (LGA) among infants delivered within 1 day of an ultrasound examination.*

Methods *Percentage differences between EFW and birth weights were calculated in 3697 pregnancies. Linear regression was used to compare the accuracy of EFW for births on each of the 6 days after an ultrasound scan with the accuracy observed among births on the same day. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value for diagnosis of SGA and LGA according to EFW was assessed.*

Results *The mean ± SD percentage difference among deliveries within 1 day of the last ultrasound scan was 0.2 ± 9.0%. Mean percentage differences were not significantly different from day 0 on days 1, 2 and 3; however, combining the data from these 4 days obscured a slight bias towards an overestimation of weight evident on day 0 and day 1. Among deliveries within 1 day of an ultrasound scan, the PPV was 61% for SGA diagnosis and 54% for LGA diagnosis.*

Conclusion *Combining data from births > 1 day after the last ultrasound examination may lead to a false conclusion that there is systematic underestimation of weight. EFW tended to underestimate the weight of macrosomic fetuses and overestimate that of small fetuses which limited sensitivity and PPV. Maternal–fetal characteristics are weak predictors of individual errors in EFW. Copyright © 2010 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Birth weight is an important predictor of perinatal morbidity and mortality[1,2], yet it cannot be measured until after birth. With the use of a combination of sonographically obtained fetal measurements of biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and femur length (FL), fetal weight can be estimated before birth in order to predict birth weight, and thereby guide the management of labor and delivery[3]. A variety of formulae have been developed for estimation of fetal weight[4–9], but none has been found to provide acceptably accurate estimates of birth weight and no single formula has emerged as superior to others[10].

In order to understand the clinical value of ultrasonographic estimates of fetal weight, validation studies documenting their accuracy and precision are required. However, existing studies comparing birth weight with estimated fetal weight (EFW) to assess the validity of EFW formulae suffer from a variety of limitations. A concern with many existing validation studies is that they included births that occurred over a wide range of days after the last ultrasound examination. Presumably, this was done out of necessity for obtaining an adequate sample size. Many studies have included births up to 1 week beyond the last ultrasound examination[11–13]. Combining the data available for infants born on the same day

ORIGINAL PAPER

as the last ultrasound examination with infants born up to 1 week later may bias estimates of EFW validity because the fetus grows 20 g per day, on average, during late pregnancy[14]. Previous studies have failed to examine the effect that including such wide ultrasound-to-delivery intervals might have on the study results.

In addition, previous studies examining the accuracy of EFW may have underestimated the amount of error that occurs in actual clinical practice. Many authors reported that ultrasound examinations were carried out exclusively by senior ultrasonographers in their studies[15–17], whereas in clinical practice those operating ultrasound machines vary in experience. Research study protocols also often call for greater repetition of measurements than would typically be obtained in clinical practice. For example, Anderson *et al.* reported a study in which two independent sonographers reviewed hard copy images from each ultrasound scan and both recorded their most accurate measurements[11]. Another study included multiple views of the abdomen, and between four and eight measurements of the fetal abdominal diameter were taken, with the average diameter used to estimate the AC[8].

Most previous validation studies have been based on small sample sizes. In a systematic review of EFW validation studies, Dudley identified 14 studies that examined what was defined in the review as a 'normal clinical population'[10]. Although these studies included the range of birth weights usually encountered in clinical practice, the median sample size was 165, and only one study had a sample size over 500. Many validation studies have focused their analyses on the subset of fetuses that are of greatest clinical interest: those that are small-for-gestational age (SGA) or large-for-gestational age (LGA). This creates an even greater restriction in the study sample and so limits the sample size even further. Among 11 studies in Dudley's review that focused exclusively on populations with a low birth weight, the median sample size was 69. In the six studies identified that focused exclusively on populations with a high birth weight, the median sample size was 26. There was, however, one large study by Miller *et al.*[12] that included 150 LGA fetuses.

In the study described herein, we sought to overcome the limitations of previous validation studies. Our specific objectives were to explore the effects of ultrasound-to-delivery interval and maternal/fetal characteristics on the distribution of measurement error in EFW in a large clinical population, and to determine the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of EFW for diagnosis of SGA and LGA among infants delivered within 1 day of their last ultrasound examination.

METHODS

We utilized a clinically detailed computerized obstetric and neonatal database (McGill Obstetrical and Neonatal Data, MOND) that has been maintained since 1 January 1978 and contains information about all infants > 500 g

delivered in the perinatal tertiary care center at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada[18]. We focused exclusively on data from births between 1996 and 2006. During this period, a routine 32-week ultrasound examination was standard practice for pregnancies documented in the database.

Six ultrasound machines are used at the institution: three Acuson Sequoias (Siemens Medical Solutions USA, Inc., Malvern, PA, USA), two Toshiba Aplios (Toshiba America Medical Systems, Inc., Tustin, CA, USA) and one GE Voluson (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). An Aloka SSD-2000 (Aloka America, Wallingford, CT, USA) was used in the early study period and newer machines were acquired from around 1999. Ultrasound examinations were performed by certified ultrasound technologists with subspecialty training in obstetric ultrasound imaging or physicians in maternal and fetal medicine with fellowship training in obstetric ultrasound examination.

EFW was calculated according to the formula by Hadlock *et al.* that incorporates four anthropometric parameters and is currently used widely in clinical practice ($\log_{10}$ weight $= 1.3596 - 0.00386$ AC $\times$ FL $+ 0.0064$ HC $+ 0.00061$ BPD $\times$ AC $+ 0.0424$ AC $+ 0.174$ FL)[7]. The measurements of HC and BPD were taken according to procedures detailed by Doubilet and Greenes[19]. AC and FL measurements were taken according to internal protocols (R. Brown, pers. comm.). Measurement of the AC is taken at the skin level from a true transverse view of the fetal abdomen at the level of the junction of the umbilical vein, portal sinus and fetal stomach (when visible). The outline of the abdomen should be as circular as possible. The measurement can be taken either as a direct circumference measurement or interpolated from anteroposterior and transverse diameter measurements. FL is measured with the beam of insonation perpendicular to the long axis of the femoral shaft, excluding the distal femoral epiphysis. The diaphysis is measured from the greater trochanter above to the lateral condyle below. The outer border of the femur is straight and the inner border is curved normally.

Our study sample consisted of singleton live births born within 1 week of the last documented estimate of fetal weight by ultrasound examination. Twin births were excluded because we were uncertain about the reliability of linking twin assignment at ultrasound imaging with twin assignment at birth. Furthermore, measurement error has been shown to vary within pairs of twins[17]. The sample was restricted to births in which the estimate of gestational age was based on the mother's last normal menstrual period (LNMP) (differentiated from spotting or other abnormal bleeding that may have occurred) and confirmed by an ultrasound examination obtained before 20 weeks where there was < 10 days discordance between menstrual date and ultrasound date. Ultrasound dates were obtained using the crown–rump length up to 13 weeks, whereas BPD was used to date the pregnancy after the first trimester[19]. Births with congenital

abnormalities and pregnancies with invalid or implausible estimates of fetal weight were excluded

We compared our study sample to similar births documented in the MOND database (singleton, live births with a confirmed gestational age and without congenital anomalies) delivered more than 6 days after the last recorded ultrasound examination using *t*-tests for means or proportions where appropriate at the α = 0.05 level.

The accuracy of fetal weight estimation was examined by calculating the mean percentage difference using the formula (EFW − birth weight)/birth weight × 100%. Therefore, a negative mean percentage difference indicates that EFW on average underestimated the actual birth weight, whereas positive values indicate overestimation. The precision of fetal weight estimation was examined by calculating the SD of the percentage differences and by examining the distribution of the absolute values of the percentage differences. The mean percentage difference allows us to estimate the systematic error and the SD gives us information about the extent of random measurement error. The absolute difference in grams was not used because it is affected by the mean birth weight of the study sample. This limits comparability with other validation studies and limits the application of study findings to the clinical setting where there is interest in a wide range of possible birth weights.

The accuracy of fetal weight estimates was compared between births within each of the included ultrasound-to-delivery intervals from 0 to 6 days. An ultrasound-to-delivery interval of 0 days indicates that the birth occurred on the same day as the last ultrasound examination; intervals of 1 to 6 days denote births 1 to 6 days, respectively after the last ultrasound examination. A linear regression model with indicator variables corresponding to each of the 6 days following ultrasound examination was used to assess differences in the accuracy of weight predictions compared with the accuracy among births on day 0. Our analysis corresponds to the model: mean percentage difference = β0 [intercept = mean percentage difference on day 0] + β1 (interval = 1) + β2 (interval = 2) + β3 (interval = 3) + β4 (interval = 4) + β5 (interval = 5) + β6 (interval = 6). We used *t*-tests to determine whether the mean percentage difference within each ultrasound-to-delivery interval and the overall mean percentage difference were significantly different from 0%.

To assess maternal and fetal determinants of the accuracy of fetal weight predictions, percentage differences and absolute percentage differences were compared between infants by sex, and various groupings of birth weight, gestational age and length at birth. Other factors examined included maternal hypertension (pre-existing or pregnancy induced), diabetes (pre-existing or pregnancy induced), obesity (defined as a body mass index (BMI) $\geq$ 30 kg/m^2) and fetal presentation. Birth date was examined to determine whether the acquisition of newer ultrasound technology improved the prediction of birth weight.

Because obesity as well as hypertension and diabetes are known to affect fetal growth, multiple linear regression was used to assess whether these comorbidities predicted mean percentage difference, independent of any effects mediated through birth weight. The effect of gestational age and the length of the fetus were not included in the model as they were highly correlated with birth weight, making it difficult to determine the independent effects of each.

Whether or not an EFW correctly identifies an SGA or LGA fetus as such is an important consideration for clinicians using EFW to inform delivery decisions. In order to determine the value of ultrasound examination data for decision making, sensitivity, specificity, PPV and NPV were calculated for identification of SGA and LGA fetuses. Likelihood ratios were also calculated in order to determine the post-test probability of SGA and LGA as well as non-SGA and non-LGA. The post-test probability was established by first calculating the post-test odds (= pretest odds (i.e. the prevalence) × likelihood ratio), then converted from odds to probability using the formula: probability = odds/(1 + odds). These analyses were restricted to births that occurred on the same day as the last ultrasound examination or the following day. Fetuses with an EFW below the 10th percentile of birth weight for gestational age[19] would be classified as SGA according to ultrasound imaging. Fetuses with an EFW above the 90th percentile for gestational age would be classified as LGA according to ultrasound examination. These findings were compared with the findings of SGA or LGA at birth based on the observed birth weight for gestational age percentile.

Statistical analysis was carried out with STATA 10 (StataCorp, College Station, TX, USA).

RESULTS

There were 6406 births within 1 week of the last documented ultrasound examination in the MOND database. Among these, there were 62 stillbirths, 722 multiple births, 541 births with congenital anomalies, 1335 births where gestational age was not based on the LNMP confirmed by early ultrasound examination, and 44 births where no EFW was recorded from the last ultrasound examination. Furthermore, we observed that five subjects had a discrepancy between EFW and birth weight that was greater than 70% of the birth weight. These were excluded because the last EFW was implausible based on visual inspection and was probably recorded incorrectly in the database. The births took place between 35 and 41 weeks' gestation and the EFWs were 100, 155, 356, 628 and 694 g. The actual birth weights were 3550, 1710, 3280, 2590 and 3355 g, respectively, which resulted in percentage differences from −76% to −97%. The remaining study sample consisted of 3697 births.

We compared our study sample with similar births documented in the MOND database (singleton, live births with a confirmed gestational age and without

congenital anomalies) delivered more than 6 days after the last recorded ultrasound examination (Table 1). The comparisons revealed statistically significant differences between the group means or proportions at the $P < 0.01$ level, with the exception of maternal age, gestational age at birth and sex. Although many of these statistically significant differences were not clinically meaningful, study mothers were more likely to have hypertension, diabetes, to be having their first child and to have a higher average BMI. Infants in the study were more likely to have been delivered by a Cesarean section, a higher percentage of the fetuses were in a breech position and a much higher percentage were induced.

Table 2 shows the effect of the ultrasound-to-delivery interval on the accuracy of weight predictions. An estimate of the mean percentage difference for all births within a week of the last ultrasound examination revealed statistically significant underestimation of birth weight by EFW. However, fetal weights among births on the same day as the last ultrasound examination were slightly overestimated (but these differences were not statistically significant). The estimated error appeared to be greater as the included ultrasound-to-delivery interval widened. Over time, birth weight became a less accurate proxy for fetal weight owing to continued fetal growth occurring between the time of the ultrasound examination and

delivery. Hence, our estimates of the extent and direction of systematic error in EFW were affected by the width of the ultrasound-to-delivery interval included in the analyses.

Mean percentage differences between the EFW and actual birth weights were not significantly different from the mean percentage difference on day 0 until day 4; however, the mean percentage difference based on combined data from observations on all four of these days obscured the overestimation observed on days 0 and 1, which was counterbalanced by negative mean percentage differences on days 2 and 3.

Although the findings in Table 2 demonstrate that the validity of EFW is affected by the ultrasound-to-delivery interval, there is evidence to suggest that the effect on precision is minimal. When births with an ultrasound-to-delivery interval of 0 or 1 day were compared with those with intervals of 2–6 days, 1.5% more births had an EFW within 15% of the actual birth weight on days 0 and 1, and most of these fell within 5% of the actual birth weight.

Table 3 shows the mean percentage differences and the percentage of births with an EFW within 5%, 10% and 15% of the actual birth weight according to various maternal and fetal characteristics. Table 3 is provided for descriptive purposes. It describes the data that are subsequently evaluated according to a multivariate linear model to determine the independent effects of each of these characteristics. It also describes how these characteristics were stratified or dichotomized for the linear model. This table was restricted to births on the same day as the last ultrasound examination or the following day. Infants ≥ 4500 g at birth contained the lowest percentage of EFW within 5% of actual birth weight. Infants born to obese mothers had the second lowest percentage of estimated weights within 5% of actual birth weight. This may mean that these groups suffer from more random errors in EFW than others.

Although we did not observe a strong effect of low birth weight (< 2500 g), the effect became apparent when we examined weight groups containing progressively smaller infants. Among the 66 deliveries of infants weighing < 2000 g, the mean $\pm$ SD percentage difference was $2.5 \pm 12.0\%$, among the 36 births < 1500 g, the mean percentage difference was $3.6 \pm 11.7\%$, and among the 10 births < 1000 g, the mean percentage difference was $10.2 \pm 13.6\%$.

According to the multivariate analysis of the association between maternal and fetal characteristics and mean percentage difference, birth weight ≥ 4500 g and presence of maternal obesity each had independent effects on the mean percentage difference, but in opposite directions. Birth weight ≥ 4500 g was associated with a mean percentage difference that was on average 5.8% less than the mean percentage difference among birth weights 3000–3499 g (95% CI, -11.2 to -0.5%). Obesity was associated with a mean percentage difference that was on average 4.2% (95% CI, 1.6–6.8%) greater than the mean percentage difference among women with a

Table 1 Characteristics of the study sample of 3697 singleton births at the Royal Victoria Hospital in Montreal, Canada with an estimated fetal weight within 1 week of delivery compared with similar births in the source population delivered more than 1 week after an ultrasound examination

Characteristic	Study sample (n = 3697)	Source population (n = 16 537)
Maternal		
Age (years)	31.3 ± 5.1	31.4 ± 4.8
Parity*		
0	1975 (53.4)†	7691 (46.5)
1	1167 (31.6)	6235 (37.7)
2	357 (9.7)	1893 (11.4)
≥ 3	197 (5.3)	717 (4.3)
Prepregnancy BMI (kg/m²)*	24.4 ± 5.2†	23.9 ± 4.9
Hypertension	260 (7.0)†	777 (4.7)
Diabetes mellitus	428 (11.6)†	1121 (6.8)
Fetal		
Birth weight (g)	3330 ± 690	3435 ± 487
GA at birth (days)	275.2 ± 17.7	275.6 ± 10.30
Male gender	1848 (50.0)	8403 (50.8)
Ultrasound-to-delivery interval (days)	2.9 ± 1.9	37.5 ± 25.2
Induced labor	1640 (44.4)†	4453 (26.9)
Cesarean delivery	993 (26.9)†	3664 (22.2)
Breech presentation*	192 (5.2)†	607 (3.7)

Values are mean ± SD or *n* (%). *Study sample: missing data for one observation on parity, 2084 observations on body mass index (BMI) and four observations on presentation; source population: missing data for one observation on parity, 9662 observations on BMI and 25 observations on presentation. †Statistically significant difference between the study sample and the source population. GA, gestational age.

Table 2 Effect of ultrasound-to-delivery interval on the accuracy of estimated fetal weight

Ultrasound-to-delivery interval (days)	n	Mean ± SD difference (%)	Difference from 0 days (95% CI)
0	271	0.49 ± 9.89	Reference group
1	828	0.09 ± 8.73	−0.40 (−1.64 to 0.85)
2	636	−0.22 ± 9.35	−0.71 (−2.00 to 0.58)
3	575	−0.43 ± 8.79	−0.91 (−2.22 to 0.40)
4	508	−1.33 ± 8.77*	−1.82 (−3.15 to −0.48)†
5	414	−1.13 ± 9.35*	−1.62 (−3.01 to −0.23)†
6	465	−2.95 ± 9.13*	−3.44 (−4.80 to −2.08)†
Overall	3697	−0.73 ± 9.11*	

*Mean percentage difference significantly different from 0 ($P < 0.05$); †Mean percentage difference significantly different from mean difference at 0 days.

Table 3 Unadjusted effect of maternal and fetal characteristics on estimated fetal weight accuracy

Characteristic	n	Mean ± SD difference (%)	Absolute difference* ≤ 5%	≤ 10%	≤ 15%
Overall	1099	0.19 ± 9.03	44.4	76.3	91.8
Birth date					
1 Jan 1996 to 31 Dec 1999	500	−0.75 ± 8.85	44.2	76.6	93.4
1 Jan 2000 to 31 Dec 2006	599	0.97 ± 9.11	44.6	76.0	90.5
Gender					
Male	564	−0.77 ± 9.02	45.9	75.0	91.3
Female	535	1.19 ± 8.94	42.8	77.6	92.3
Birth weight (g)†					
< 2500	135	0.57 ± 10.22	41.3	73.2	89.1
2500–2999	211	1.70 ± 8.95	40.8	76.8	93.4
3000–3499	327	0.69 ± 8.56	46.5	79.2	91.4
3500–3999	303	−0.60 ± 9.02	47.2	75.3	92.4
4000–4499	96	−1.82 ± 8.46	43.8	71.9	91.7
≥ 4500	24	−4.23 ± 8.01	33.3	79.2	91.7
Gestational age at birth (weeks)					
< 35	78	0.57 ± 11.56	43.6	70.5	85.9
35–36	77	−2.10 ± 7.83	45.5	75.3	93.5
37–40	786	0.22 ± 8.78	44.2	78.0	92.4
≥ 41	158	0.97 ± 9.30	45.6	70.9	91.1
Hypertension					
Yes	79	2.04 ± 8.74	41.8	81.0	93.7
No	1020	0.05 ± 9.04	44.6	75.9	91.7
Diabetes					
Yes	108	−0.62 ± 8.97	38.0	77.8	93.0
No	991	0.28 ± 9.04	45.1	76.1	91.7
Obese (BMI ≥ 30 kg/m^2)†					
Yes	59	2.93 ± 10.85	37.3	71.2	84.8
No	388	−0.84 ± 9.06	45.4	75.5	92.8
Neonatal length (cm)†					
≤ 49	341	0.86 ± 9.59	42.2	73.9	90.3
> 49–51	288	0.98 ± 9.90	47.9	78.8	90.6
> 51–52.5	192	−0.41 ± 9.03	45.8	77.1	90.6
> 52.5	207	−1.51 ± 8.31	46.9	76.3	95.2
Presentation					
Breech	68	−0.55 ± 11.31	44.1	69.1	85.3
Other	1031	0.24 ± 8.86	44.4	76.7	92.2

*Percentage of births with estimated fetal weight within 5%, 10% and 15% of birth weight. †Missing data for birth weight, BMI and neonatal length. BMI, body mass index.

BMI < 30 kg/m^2. The statistical model incorporated the obesity variable at the expense of 652 observations with missing data for height, prepregnancy weight or both. The effects of the other potential determinants, with the exception of sex, however, remained the same in a model without obesity that incorporated the full dataset (results available on request). In the model without obesity, female sex was associated with a mean percentage difference that was on average 1.6% (95% CI, 0.5–2.7%) greater than the mean percentage difference among males. The

model explained only a small amount of the variation in percentage differences (adjusted $R^2 = 3.5\%$).

Table 4 shows the data used for calculation of sensitivity, specificity, PPV and NPV for detection of SGA fetuses. Sensitivity, the true-positive rate, was 69%. Specificity, the true-negative rate, was 93%. The PPV was 61% and the NPV was 95%. The post-test probability of SGA was 48% vs. a pretest probability of 10%, which is the prevalence of SGA in the population (likelihood ratio, 9.4). The post-test probability of non-SGA was 21% vs. a pretest probability of 90% (likelihood ratio, 0.3).

The data used for calculation of sensitivity, specificity, PPV and NPV for detection of LGA fetuses are shown in Table 5. Sensitivity was 68% and specificity was 94%. PPV was 54% and NPV was 96%. The post-test probability of LGA was 53% vs. a pretest probability of 10% based on the prevalence of LGA (likelihood ratio, 11.2). The post-test probability of non-LGA was 21% *versus* a pretest probability of 90% (likelihood ratio, 0.3).

DISCUSSION

This study is the first to utilize a large clinical database of births delivered within 1 week of the last documented ultrasound examination to determine the effect of ultrasound-to-delivery interval on the validity of EFW. Our results also provide a unique picture of the validity of EFW in routine clinical practice, rather than the research setting. Moreover, the large sample size and diversity in patient types obtained in our tertiary-care setting allowed us to further explore the measurement error in EFW according to maternal and fetal characteristics.

We demonstrated that the inclusion of wide ultrasound-to-delivery intervals in validation studies, as has been routinely done in past studies, introduces bias to estimates of measurement error. We found that estimates of fetal weight were on average slightly higher than the actual birth weights recorded later that day. As time between the ultrasound examination and the subsequent delivery increased, growth of the fetus caused increasingly negative discrepancies between EFW and birth weight. This could lead to a false conclusion that there is systematic underestimation of weight by ultrasound imaging when in fact there may be a slight positive bias. In general, inclusion of intervals of 0–1 days will introduce only minimal bias into estimates of measurement error, whereas including births more than 1 or 2 days after the last ultrasound examination will incorporate more error introduced by continued fetal growth.

With a mean percentage difference close to 0% and a SD of 9%, our results are consistent with the findings of previous studies[7,15,20,21]. Most validation studies have failed to demonstrate substantial systematic bias in weight estimates obtained with one of the Hadlock formulae among a sample with a wide range of birth weights. The SD observed in our study falls in the middle of the range of reported variations in accuracy of fetal weight estimation. Some studies have observed SDs in percentage differences as low as 7–8%[7,20], whereas others have found larger variations. For example, Simon *et al.* noted a SD of 11.7% among estimates obtained with the Hadlock formula based on four measurements[15].

We observed that the most reliable EFWs were obtained among deliveries with birth weights of 3000–3999 g. Consistent with previous studies[7,12,15], we found a tendency to overestimate the weight of small fetuses (< 3000 g) and underestimate the weight of large fetuses (≥ 4000 g), with the effect most pronounced for fetuses ≥ 4500 g. Hadlock and others have found the effects to be of similar magnitude in either direction[7,12,15]. The results of Kurmanavicius *et al.*, however, were consistent with ours[13]; they found percentage errors close to zero below 3500 g and increasingly negative mean percentage errors beyond 3500 g. Interestingly, when we examined progressively smaller birth weight groups in our study, we observed a trend toward increasing overestimation.

Four different maternal/fetal characteristics were associated with unusually high variation in the percentage errors: breech presentation (SD, 11.3), gestational age < 35 weeks (SD, 11.6), birth weight < 2500 g (SD, 10.2) and maternal obesity (SD, 10.8). According to our multivariate models, high birth weight and maternal obesity have independent effects on the accuracy of EFW. Most importantly, however, the regression model with all maternal/fetal characteristics combined explained only a very small percentage of the total variation in percentage differences. This suggests that most of the error in EFW is

Table 4 Agreement between small-for-gestational age (SGA) diagnoses based on estimated fetal weight (EFW) determined by ultrasound imaging and on birth weight among births within 1 day of the last ultrasound examination

Diagnosis based on EFW	Diagnosis based on birth weight		
	Non-SGA	*SGA*	*Total*
Non-SGA	872 (79.3)	49 (4.5)	921
SGA	69 (6.3)	109 (9.9)	178
Total	941	158	1099

Data are given as *n* (%) or *n*.

Table 5 Agreement between large-for-gestational age (LGA) diagnoses based on estimated fetal weight (EFW) determined by ultrasound imaging and on birth weight among births within 1 day of the last ultrasound examination

Diagnosis based on EFW	Diagnosis based on birth weight		
	Non-LGA	*LGA*	*Total*
Non-LGA	934 (85.0)	34 (3.1)	968
LGA	60 (5.5)	71 (6.5)	131
Total	994	105	1099

Data are given as *n* (%) or *n*.

unrelated to measured maternal or fetal characteristics. It is therefore unlikely that clinicians will be able to predict which individual fetal weights will be underestimated or overestimated based such characteristics.

Sensitivities for detection of SGA or LGA are less than optimal because prediction of birth weight is less accurate for the low and high ends of the birth-weight spectrum. According to a review by Chang *et al.*, the sensitivity of an SGA diagnosis has been reported to range from 33% to 89%[22]. The potential for variability in these estimates is high, however, according to rate of physician intervention and the width of the ultrasound-to-delivery interval in each study.

The likelihood ratios for a positive SGA or LGA diagnosis by ultrasound examination are around 10, which is considered to be strong evidence for the usefulness of a test[23,24]. After an ultrasound examination has identified a fetus as SGA or LGA, the probability of SGA or LGA increases from 10% to 50–60%. This indicates that an EFW diagnosis of SGA or LGA can provide important information; however, this should be utilized in the context of other clinical indicators of SGA or LGA.

A concern with these data is that they constitute a highly selected sample, although it seems unlikely that this would introduce bias into the results of the distribution of measurement error. The high intervention rate in this sample, however, may influence the results of the sensitivity and specificity analysis. Fetuses classified as SGA or LGA according to an ultrasound examination will be over-represented among births within 1 day of an ultrasound examination owing to increased physician intervention. Sensitivity will tend to be overestimated as some SGA fetuses classified as non-SGA by ultrasound imaging will not be delivered as hastily (and therefore not within 1 day of the last ultrasound examination, and hence excluded from our analysis). Specificity will tend to be underestimated as fetuses classified as SGA or LGA according to ultrasound examination findings will be over-represented.

In conclusion, restricting study samples to births within 1 day of the last ultrasound examination helps to obtain the most accurate estimate of the error in EFW. Owing to the large number of births within 1 day of an ultrasound examination, our data were uniquely suited to assessing the extent and direction of measurement error in EFW without introducing bias resulting from continued fetal growth. This study provides more accurate estimates of measurement error in EFW, and the sensitivity, specificity, PPV and NPV of EFW for diagnosis of SGA and LGA. We highlight the importance of improving prediction of birth weight in the tails of the birth weight continuum to improve prenatal diagnoses of SGA and LGA. Finally, we demonstrate that the majority of the error in EFW is unexplained by maternal and fetal characteristics.

ACKNOWLEDGMENTS

REFERENCES

1. Schlesinger ER, Allaway NC. The combined effect of birth weight and length of gestation on neonatal mortality among single premature babies. *Pediatrics* 1955; **15**: 698–704.
2. McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. *N Engl J Med* 1999; **340**: 1234–1238.
3. Chatfield J. ACOG issues guidelines on fetal macrosomia. *Am Fam Physician* 2001; **64**: 169–170.
4. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br J Obstet Gynaecol* 1975; **82**: 689–697.
5. 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.
6. 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.
7. 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.
8. Rose BI, McCallum WD. A simplified method for estimating fetal weight using ultrasound measurements. *Obstet Gynecol* 1987; **69**: 671–675.
9. Combs CA, Jaekle RK, Rosenn B, Pope M, Miodovnik M, Siddiqi TA. Sonographic estimation of fetal weight based on a model of fetal volume. *Obstet Gynecol* 1993; **82**: 365–370.
10. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 2005; **25**: 80–89.
11. Anderson NG, Jolley IJ, Wells JE. Sonographic estimation of fetal weight: comparison of bias, precision and consistency using 12 different formulae. *Ultrasound Obstet Gynecol* 2007; **30**: 173–179.
12. Miller JMJ, Kissling GA, Brown HL, Gabert HA. Estimated fetal weight: Applicability to small- and large-for-gestational-age fetus. *J Clin Ultrasound* 1988; **16**: 95–97.
13. Kurmanavicius J, Burkhardt T, Wisser J, Huch R. Ultrasonographic fetal weight estimation: accuracy of formulas and accuracy of examiners by birth weight from 500 to 5000 g. *J Perinat Med* 2004; **32**: 155–161.
14. 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.
15. Simon NV, Levisky JS, Shearer DM, O'Lear MS, Flood JT. Influence of fetal growth patterns on sonographic estimation of fetal weight. *J Clin Ultrasound* 1987; **15**: 376–383.
16. Ben-Haroush A, Yogev Y, Mashiach R, Hod M, Meizner I. Accuracy of sonographically estimated fetal weight in 840 women with different pregnancy complications prior to induction of labor. *Ultrasound Obstet Gynecol* 2004; **23**: 172–176.
17. Danon D, Melamed N, Bardin R, Meizner I. Accuracy of ultrasonographic fetal weight estimation in twin pregnancies. *Obstet Gynecol* 2008; **112**: 759–764.

18. Smith LP, DeLeon A, Funnell WR, Lalonde AB, McLean FH, Usher RH. A research oriented system for McGill Obstetrical and Neonatal Data (MOND). *Acta Obstet Gynecol Scand* 1982; **102** (Suppl): 49–50.
19. Doubilet PM, Greenes RA. Improved prediction of gestational age from fetal head measurements. *AJR Am J Roentgenol* 1984; **142**: 797–800.
20. Shamley KT, Landon MB. Accuracy and modifying factors for ultrasonographic determination of fetal weight at term. *Obstet Gynecol* 1994; **84**: 926–930.
21. Smulian JC, Ranzini AC, Ananth CV, Rosenberg JC, Vintzileos AM. Comparison of three sonographic circumference measurement techniques to predict birth weight. *Obstet Gynecol* 1999; **93**: 692–696.
22. Chang TC, Robson SC, Boys RJ, Spencer AD. Prediction of the small for gestational age infant: which ultrasonic measurement is best? *Obstet Gynecol* 1992; **80**: 1030–1038.
23. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 389–391.
24. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 703–707.