Published in final edited form as:

Paediatr Perinat Epidemiol. 2012 May ; 26(3): 199–207. doi:10.1111/j.1365-3016.2012.01261.x.

Predicting Large Fetuses at Birth: Do multiple ultrasound examinations and longitudinal statistical modelling improve prediction?

Jun Zhang^{1,*}, Sungduk Kim^{2,*}, Jagteshwar Grewal², and Paul S. Albert²

¹MOE and Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine and School of Public Health, Shanghai, P. R. China

²Division of Epidemiology, Statistics and Prevention Research, Eunice Shriver Kennedy National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, U.S.A

SUMMARY

Predicting large fetuses at birth has long been a challenge in obstetric practice. We examined whether ultrasound examinations at multiple times during pregnancy improve accuracy of prediction using repeated, longitudinal statistical modelling; and whether adding maternal characteristics improves the accuracy of prediction. We used data from a previous study conducted in Norway and Sweden from 1986 – 1989 in which each pregnant woman had four ultrasound exams at around 17, 25, 33 and 37 weeks of gestation. At birth, infant size was classified as large-for-gestational-age (LGA, > 90th centile) and macrosomia (>4,000 g) or not. We used a longitudinal random effects model with quadratic fixed and random effects to predict term LGA and macrosomia at birth. Receiver-operator curves and mean-squared error were used to measure accuracy of the prediction. Ultrasound exam around 37 weeks had the best accuracy in predicting LGA and macrosomia at birth. Adding multiple ultrasound exams in earlier gestation did not improve the accuracy. Adjusting for maternal characteristics had limited impact on the accuracy of prediction. Thus, a single ultrasound exam at late gestation close to birth is the simplest method currently available to predict LGA and macrosomia.

Keywords

birthweight; fetal weight; large-for-gestational-age; macrosomia; prediction

INTRODUCTION

A large fetus or macrosomia is a major concern to obstetricians, affecting both maternal and fetal morbidity and mortality.¹ For instance, shoulder dystocia can be devastating to both physician and patient and has a higher risk of fetal mortality. Thus, accurate estimation and prediction of birthweight has long been a quest for researchers and obstetricians. It is particularly important given the epidemic of maternal obesity.

Corresponding author: Dr. Paul Albert, Biostatistics & Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Shriver Kennedy National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, U.S.A., Tel: 301 496-5582, Fax: 301-402-2084.

*J Zhang and S Kim contributed equally to the study.

However, estimating fetal weight by ultrasound at one time point suffers from inaccuracy with an error of 10 – 15%.² For example, if a true fetal weight is 4,000 grams, the estimated fetal weight (EFW) could vary from 3,400 to 4,600 grams. Peregrine et al.³ showed that the sensitivity of ultrasound estimate of fetal weight > 4000 g was less than 50%. Bear in mind that most previous studies were carefully done by highly trained investigators. Errors in ultrasound measurements are likely to be even greater in daily practice.

In addition to measurement errors, the inaccuracy of estimating fetal weight can also come from formulae that are used to calculate the fetal weight. Most formulae are based on a linear regression of 2 dimensional measurements of fetal biometry (e.g., head circumference (HC), abdominal circumference (AC) and femur length (FL)).⁴ Studies have shown that most formulae have a tendency to underestimate fetal weight in large fetuses and overestimate fetal weight in small fetuses², which may be related to regression to the mean. Another source of error in estimating fetal weight could be related to using one- or two-dimensional measures to estimate EFW, which is a mass composed of a three-dimensional volume $\times$ density.

Over the years, investigators have tried to predict macrosomia prospectively with single or multiple ultrasound examinations weeks before birth. The general conclusion was that the last ultrasound close to birth is most predictive.^{5–10} However, previous studies used simple statistical methods without taking advantage of repeated, longitudinal nature of the data. We apply a newly developed statistical modelling approach which allows us to predict macrosomia or large for gestational age (LGA) at birth with as few as one single ultrasound measurement. We then use the statistical methodology to address four questions: (1) Do ultrasound examinations at multiple times during pregnancy improve accuracy of prediction of large infants using repeated, longitudinal statistical modelling? (2) If it is so, how many times should the examinations be done? (3) Will adding covariates, i.e., maternal characteristics, improve the accuracy of prediction? (4) Instead of using EFW to predict birthweight, will simultaneous modelling of individual 2-dimensional foetal biometry measures actually reduce errors of the estimate and improve the accuracy of prediction?

METHODS

The “Successive Small-for-Gestational Age Births” study (SGA study) was initiated and sponsored by the National Institute of Child Health and Human Development, National Institutes of Health in the U.S.A. Its purpose was to describe and characterize the different types of intra-uterine growth restriction and their consequences, to assess the associated risk factors, and to provide a basis for preventive measures. Starting in 1986, it ran concurrently in Bergen and Trondheim (Norway) and Uppsala (Sweden), and University of Alabama at Birmingham (USA) recruiting pregnant para 1 and 2 mothers at high risks of having an SGA birth and a random (control) sample of the total pregnant population. Data collection ended in late 1989, when the last-born children have reached 13 months of age.¹¹ Women invited to participate in the study were given an informed consent form, which was discussed and signed prior to joining the study. The study protocol was also reviewed and approved by the respective local ethical committees.

The current study used the Scandinavian data only because women in the Scandinavian cohort differed substantially from women in the Alabama cohort. More importantly, women were more likely to have completed all follow-up visits on time in the Scandinavian cohort. Eligible for enrollment in the Scandinavia part of the study were parous women of Caucasian origin who spoke one of the Scandinavian languages, had a singleton pregnancy, and were registered by the study center prior to the 20th gestational week. Women were considered as being at high risk in the event they had any of the following 11 risk factors:

previous delivery of a low birth weight infant (i.e. below 2750 g), previous stillbirth or neonatal death, history of two or more spontaneous miscarriages, history of phlebitis, initial systolic blood pressure above 140 mmHg, a previous preterm birth, prepregnancy weight or weight at the first clinic visit of less than 50 kg, currently smoking or smoked at conception, and use of alcohol. Of the 6,354 women referred to the study, 5,722 were eligible for participation and made an appointment. Of these women, 1,384 were classified as high risk for IUGR and enrolled for the study. In addition, a random sample of 561 women from a Nordic population of Caucasian women was selected as a reference group. Thus, a total of 1,945 subjects were recruited in the study. A detailed description of the study was published elsewhere.¹¹

At entry detailed maternal demographic characteristics and medical history were collected. Four prenatal study visits were scheduled at approximately 17, 25, 33 and 37 weeks of gestation. At each visit, woman's information was updated and an ultrasound exam was performed for fetal biometry. At birth, infant weight was carefully measured and medical information was extracted from medical charts.

Among the 1,945 women enrolled in the study, 393 (20.2%) failed to complete more than one of the four study examinations. Most of them, 296 or 75.3% gave 'social inconvenience' as the reason for not continuing with the study. The women who had incomplete measurements were compared with those who completed the scheduled visits with respect to all socio-demographic and medical variables available and no significant differences were found. For the purpose of this study, we further restricted to women who have all four ultrasound examinations (N = 1,023).

Among the ultrasound parameters, three fetal measures were used in the current study: biparital diameter (BPD), middle abdominal diameter (MAD) and femur length (FL). Abdominal circumference (AC) was calculated as $3.1416 \times MAD$. Based on these parameters, fetal weight was estimated according to the following formula by Hadlock et al.⁴: $\text{Log}_{10} \text{EFW} = 1.335 - 0.0034 \text{ AC} \times \text{FL} + 0.0316 \text{ BPD} + 0.0457 \text{ AC} + 0.1623 \text{ FL}$. Macrosomia is defined as birthweight > 4,000 grams. Large-for-gestational-age (LGA) is defined as birthweight above the 90th percentile at a given gestational week.¹²

Statistical methods

For a single longitudinal ultrasound measure of EFW, we built a random effect model with quadratic fixed and random effects as follow:

$$Y_{ij} = X'_{ij}\beta + \theta_0 + \theta_1 t_{ij} + \theta_2 t_{ij}^2 + b_{i0} + b_{i1} t_{ij} + b_{i2} t_{ij}^2 + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} denotes the log-transformed EFW, X_{ij} denotes a vector of fixed effect covariates such as maternal age, prepregnancy body mass index, previous history of SGA, smoking during pregnancy, β denotes a vector of corresponding regression parameters, t_{ij} is the j th gestational age for i th women, and $b_i = (b_{i0}, b_{i1}, b_{i2})'$ denotes the random effects, where, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, I$, n_i denotes the number of repeated time points for i th women, and I denotes the number of women. The parameters θ_0 , θ_1 , θ_2 and describe the overall fetal growth with a quadratic function. Further, we assume that the residual errors ε_{ij} are distributed with independent normal distributions with mean 0 and variance σ_{ε}^2 . The random effects are assumed to have mean zero and be correlated with each other.¹³ That is, we assume that the random effect $b_i = (b_{i0}, b_{i1}, b_{i2})'$ is a multivariate normal distribution with mean 0 and covariance matrix Ω , where Ω is unstructured variance-covariance matrix. The random effects are interpreted as individual departures in an individual's growth curve relative to the average fetal growth curve in the population. This random effects growth

model provides a more parsimonious model than individually estimating each fetus' longitudinal pattern, which is commonly done in the literature. Further, the use of this model allows us to more easily link together fetal growth patterns with the binary birth outcome of LGA or macrosomia.

We link the probability of LGA or macrosomia to the ultrasound measurements through predicted values of individual ultrasound measurements at 39 weeks. We correct for the measurement error associated with the prediction using a standard measurement error correction.¹⁴ Let S_i denote a binary event such as LGA or macrosomia. The probability of this binary variable is linked to the longitudinal process as

$$P(S_i|b_i)=\Phi[W_i'\gamma+\alpha(b_{i0}+b_{i1}t_*+b_{i2}t_*^2)], \quad (2)$$

where Φ denotes a cumulative density function of standard normal distribution, W_i denotes a vector of covariates for i th women, γ denotes corresponding vector of regression coefficients, and t_* is a time point near the time of birth such as 39 weeks of gestation. We note that W_i may share common components with X_{ij} in equation (1). The parameter α characterizes the relationship between the longitudinal process and the binary event. We note that the model in (2) is called a probit-link regression model in the generalized linear model framework. The joint model (1) and (2) relates the longitudinal fetal growth pattern to the probability of an abnormal birth outcome through an individual's predicted measurement at time t_* . More details on the prediction using a single longitudinal ultrasound measurement are provided in a study by Albert.¹⁵ Equations (1) and (2) describe the prediction approach for a single longitudinal measurement such as estimated fetal weight which is derived from individual fetal measurements at each follow-up time. We refer to the resulting predictor (using (1) and (2)) as the univariate predictor. Alternatively, we could directly model the longitudinal pattern for each fetal measurement (BPD, MAD, and FL) and then use them simultaneously to predict LGA or macrosomia at birth. This requires a multivariate extension of the previously discussed approach which is described in the Appendix. We refer to this predictor as the multivariate predictor.

We estimated the linear mixed model (1) for a single random longitudinal measure using standard mixed model software (lme in R package). In order to appropriately develop a predictor of LGA or macrosomia, we performed bootstrap cross-validation by dividing the whole data into two data sets randomly with a 40% and 60% split of the sample into a training and test set, respectively. The training set is used to develop the predictor while the test set data is used to validate the predictor with different accuracy measures. For both the univariate and multivariate predictors, we fit the longitudinal models and estimated the parameters of the predictor using only the training dataset with all four longitudinal measurements on each study participant. For different designs (i.e., using measurements at about 17, 25, 33, and 37 weeks as well as using subsets of these longitudinal measurements), we evaluated the diagnostic accuracy of the predictor using the test dataset. We evaluated the accuracy using two approaches. First, we used the Receiver Operator Characteristic curve (ROC), which is a standard approach for estimating the accuracy for a binary classification using a continuous marker. The ROC curves compute the area under the ROC curve (AUC), where a value of 1 corresponds to a perfect classification, while a value of 0.5 corresponds to completely random classification. The AUC can be interpreted as the probability that the predictor from a randomly selected pair of LGA and non-LGA births would correctly identify the LGA birth (i.e., the LGA has a larger predictive probability than the non-LGA birth). Additionally, we also used the mean-squared error (MSE) of prediction, which is a measure for assessing absolute risk (the average squared difference between the

predicted probability and the birth outcome). Third, we examined the frequency of correctly classifying high and low risk women with different predictors.

Although the approach does not require complete data for developing the predictor, we limited ourselves to those SGA participants with complete data (4 measurements) since we evaluated predictive accuracy across designs (different combinations of T1, T2, T3, and T4).

RESULTS

Table 1 presents the baseline characteristics of the study population. The mean maternal age was 28 years and mean prepregnancy body mass index was 21.5 kg/m². 72% had one previous birth and 61% smoked at enrollment. Nearly a quarter of them had previous SGA births but very few had diabetes or gestational diabetes. The mean gestational age at birth was 40 weeks and mean birthweight was 3562 grams. 17% of births were macrosomic and 11% were LGA.

Figure 1 illustrates the timing of ultrasound examinations and birth, which was around 17, 25, 34, 37 and 40 weeks. The mean interval between the last ultrasound exam and birth was 3 weeks (interquartile interval: 2 – 4 weeks). The three fetal growth curves represent the median, 10th and 90th percentiles of the reference established by Hadlock et al.¹⁶, which is commonly used as a fetal growth standard. This figure indicates that our study population had a fetal growth pattern similar to that of the general population.

Based on 100 bootstrap cross-validations, Table 2 shows that prediction of LGA at birth is poor in the 2nd trimester. It was substantially improved in the 3rd trimester. Ultrasound exam around 37 weeks had the best accuracy. Adding multiple ultrasound exams in earlier gestation did not improve the accuracy. Adjusting for covariates has limited impact on the accuracy of prediction. Likewise, Table 3 suggests that one ultrasound exam around 37 week of gestation will be sufficient in predicting macrosomia. Adjusting for covariates is optional. When we use three individual fetal measures to predict LGA or macrosomia using the multivariate predictor, the results were virtually the same (methods and tables are included in the appendix). As suggested by Janes et al. (2008), we also computed reclassification frequencies based on comparing risk prediction using only the T4 measurements versus using all four measurements (T1, T2, T3, T4). We classified all individuals into three risk groups for both models according to their risk as < 10%, 10%–35%, and > 35%. Median frequencies were estimated using 100 bootstrap cross-validated samples. Table 4 and Table 5 (Table A3 and Table A4) provide cross-tabulating risk predictions from the proposed model. The results are shown for the large-for-gestational age outcome with EFW as the ultrasound variable, but results were similar for the macrosomia outcome and when the three separate ultrasound variables were used. For individuals classified as low risk (<10%) with T1–T4, 97% were also classified as low risk using only T4. Likewise, for those classified as high risk (>35%) with T1–T4, close to 80% were also classified as high risk using only T4. This along with our estimates of AUC and MSE provide evidence that a single measure of EFW at T4 is sufficient for prediction.

DISCUSSION

Our prospective study in conjunction with an innovative statistical modelling approach indicates that multiple ultrasound exams during pregnancy do not improve accuracy of prediction of large infants even with longitudinal statistical modelling. Adding other maternal characteristics in the prediction model provides very limited improvement. Further, we show that a multivariate predictor using individual 2-dimensional fetal biometry

measures has a similar accuracy of prediction as EFW. Thus, a single ultrasound exam at late gestation is the simplest method currently available to predict LGA and macrosomia.

The model was based on a linear mixed model with quadratic terms for log-transformed ultrasound measurements. The model fit the data very well, and has an advantage over more complex models for fetal growth in that it can easily be fit by the practitioner. Alternative cubic spline-based approaches have recently been proposed for modelling fetal growth. These approaches could be incorporated into our predictive framework²⁰.

Consistent with our study, Larsen et al.¹⁰ found that with increasing gestational age, the accuracy of birthweight prediction improves. However, prediction based on the average of repeated weight estimates or linear extrapolation from two estimates or extrapolation by a second order polynomial fitted to four estimates did not improve accuracy compared to prediction based on the last estimate before delivery. The statistical modelling approach we took in this paper has some advantages over traditional more simple data analytic approaches for predicting LGA/macrosomia from longitudinal ultrasound measurements. First, the approach allowed for the development of a predictor using irregularly timed measurements (although measurements were targeted at 17, 25, 33, and 37 weeks of gestation, as indicated in Figure 1, there was variation in when the observations were actually taken). Second, the modelling approach allowed us to evaluate a predictor when fetal biometry was assessed at only a single time point (say at T4). We could not formulate such a predictor if we modeled all curves individually and did not incorporate a random effects modelling structure. Nahum et al.¹⁷ evaluated a number of measureable maternal and fetal characteristics to predict term birthweight. They found that combinations of gestational age, parity, fetal sex, maternal height, maternal weight and third-trimester maternal weight gain rate could explain 33% of the variance in birthweight. Term fetal macrosomia was predicted with 80% sensitivity using a prediction cutoff of 3550 g. However, such accuracy is still too low to be clinically useful. Our study found that when fetal biometry was used, maternal characteristics did not provide much additional improvement.

It is worth noting that the prevalence of maternal smoking during pregnancy in this study population was high. Given the fetal growth patterns in smokers may be different from those in nonsmokers, the generalizability of our findings may be limited to an unknown degree.

Predicting a large fetus is clinically important particularly given the obesity epidemic. The consequences of false diagnosis include a high rate of cesarean delivery (for false positive) and maternal and neonatal complications (for false negative).¹⁸ Unfortunately, repeated ultrasound measurements and advanced statistical modelling do not improve accuracy of prediction for LGA or macrosomia.

Acknowledgments

This research was supported in part by the intramural research program of Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

References

1. Esakoff TF, Cheng YW, Sparks TN, Caughey AB. The association between birthweight 4000 g or greater and perinatal outcomes in patients with and without gestational diabetes mellitus. *American Journal of Obstetrics and Gynecology*. 2009; 200:672.e1–4. [PubMed: 19376489]
2. Scioscia M, Vimercati A, Ceci O, Vicino M, Selvaggi LE. Estimation of birth weight by two dimensional ultrasonography. *Obstetrics and Gynecology*. 2008; 111:57–65. [PubMed: 18165393]

3. Peregrine E, O'Brien P, Jauniaux E. Clinical and ultrasound estimation of birth weight prior to induction of labor at term. *Ultrasound in Obstetrics and Gynecology*. 2007; 29:304–9. [PubMed: 17290365]
4. 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. *American Journal of Obstetrics and Gynecology*. 1985; 151:333–7. [PubMed: 3881966]
5. Hedrianna HL, Moore TR. A comparison of single versus multiple growth ultrasonographic examinations in predicting birth weight. *American Journal of Obstetrics and Gynecology*. 1994; 170:1600–6. [PubMed: 8203416]
6. Ben-Haroush A, Chen R, Hadar E, Hod M, Yogev Y. Accuracy of a single fetal weight estimation at 29 – 34 weeks in diabetic pregnancies: can it predict large-for-gestational-age infants at term? *American Journal of Obstetrics and Gynecology*. 2007; 197:497.e1–6. [PubMed: 17980186]
7. Ben-Haroush A, Yogev Y, Hod M, Bar J. Predictive value of a single early fetal weight estimate in normal pregnancies. *European Journal of Obstetrics Gynecology and Reproductive Biology*. 2007; 130:187–92.
8. Bryant DR, Zador I, Landwehr JB, Wolfe HM. Limited clinical utility of midtrimester fetal morphometric percentile rankings in screening for birth weight abnormalities. *American Journal of Obstetrics and Gynecology*. 1997; 177:859–63. [PubMed: 9369834]
9. Pressman EK, Bienstock JL, Blakemore KJ, Martin SA, Callan NA. Prediction of birth weight by ultrasound in the third trimester. *Obstetrics and Gynecology*. 2000; 95:502–6. [PubMed: 10725480]
10. Larsen T, Greisen G, Petersen S. Prediction of birth weight by ultrasound-estimated fetal weight: a comparison between single and repeated estimates. *European Journal of Obstetrics Gynecology and Reproductive Biology*. 1995; 60:37–40.
11. Bergsjø P, Hoffman HJ, Davis RO, Goldenberg RL, Lindmark G, Jacobsen G, Cutter G, Markestad T, Nelson KG, Bakkteteig LS. Preliminary Results from the Collaborative Alabama and Scandinavian Study of Successive Small-for-Gestational Age Births. *Acta et Obstetrics Gynecology Scand*. 1989; 68:19–25.
12. Bjerkedal T, Skjaerven R. Percentiles of birth weight and crown-heel length for single live births. *Tidsskr Nor Laegeforen*. 1980; 100:1088–91. [PubMed: 7414593]
13. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
14. Carroll RJ, Spiegelman CH, Lan KKG, Abbott RD. On errors-in-variables for binary regression models. *Biometrika*. 1984; 71:19–25.
15. Albert, PS. A linear mixed model for predicting a binary event under random effects misspecification. *Press at Statistics in Medicine*; 2011.
16. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology*. 1991; 181:129–33. [PubMed: 1887021]
17. Nahum GG, Stanislaw H, Huffaker J. Accurate prediction of term birth weight from prospectively measurable maternal characteristics. *Journal of Reproductive Medicine*. 1999; 44:705–12. [PubMed: 10483541]
18. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Annals of internal medicine*. 2008; 149:751–760. [PubMed: 19017593]
19. Melamed N, Yogev Y, Meizner I, Mashiach R, Ben-Haroush A. Sonographic prediction of fetal macrosomia: the consequences of false diagnosis. *Journal of Ultrasound in Medicine*. 2010; 29:225–30. [PubMed: 20103792]
20. Villandre L, Hutcheon JA, Perez ME, Abehnaim H, Jacobsen G, Platt RW. Modeling fetal weight for gestational age: a comparison of a flexible multi-level spline-based model with other approaches. *International Journal of Biostatistics*. 2011; 7:Article 32. [PubMed: 21931571]

APPENDIX

We extended this univariate random effects model, where we modeled the estimated fetal weight over time, to a multivariate random effects model, where we separately modeled three anthropomorphic measurements longitudinally such as biparital diameter (BPD),

middle abdominal diameter (MAD) and femur length (FL). The multivariate random effect model we used is given by

$$Y_{ijk} = X'_{ijk}\beta_k + \theta_{0k} + \theta_{1k}t_{ijk} + \theta_{2k}t_{ijk}^2 + b_{i0k} + b_{i1k}t_{ijk} + b_{i2k}t_{ijk}^2 + \varepsilon_{ijk}, \quad (3)$$

where X_{ijk} denotes a vector of fixed effect covariates, β_k denotes corresponding regression parameters, and $b_{ik} = (b_{i0k}, b_{i1k}, b_{i2k})'$ denotes the random effects for k th ultrasound measurement, where $i = 1, 2, \dots, I, j = 1, 2, \dots, n_{ik}, k = 1, 2, \dots, K, n_{ik}$ denotes the number of repeated time points for k th ultrasound measurement of i th women. Further, we assume that the residual error $\varepsilon_{ij} = (\varepsilon_{ij1}, \varepsilon_{ij2}, \dots, \varepsilon_{ijK})'$ is distributed with multivariate normal distribution with mean 0 and covariance matrix Σ_ε . Further, we assume that the random effect is a multivariate normal distribution with mean 0 and covariance matrix Ω , where Ω is unstructured variance-covariance matrix. As in the univariate case, we used the multivariate model to predict these measurements at 39 weeks as

$$P(S_i|b_i) = \Phi[W'_i\gamma + \sum_k \alpha_k(b_{i0k} + b_{i1k}t_{ik} + b_{i2k}t_{ik}^2)]. \quad (4)$$

We estimated the linear mixed model for a single random longitudinal measurement using standard mixed model software (lme in R package). We used the method of Doran and Lockwood (2006) to fit the multivariate linear mixed model (3).

Doran HC and Lockwood JR. Fitting Value-Added Models in R. *Journal of Educational and Behavioral Statistics*. 2006;31: 205–230.

Figure 1. Estimated fetal weight dot and birthweight circle by gestational age in the study population. The fetal growth curves (representing the 10th, 50th and 90th percentiles) are based on the normal reference by Hadlock et al.¹⁶

Table 1

Baseline characteristics of study population

Variables	Cohort (N = 1023) N (%)
Maternal age (mean, years) (SD)	28.1 (4.1)
Prepregnancy body mass index (mean, kg/m ²) (SD)	21.5 (3.1)
Parity (%)	
1	735 (71.9)
Smoking at enrollment (%) (cig./day)	
0	402 (39.3)
1–9	168 (16.4)
10–19	362 (35.4)
20+	91 (8.9)
History of small-for-gestational-age birth (%)	242 (23.7)
Existing diabetes (%)	3 (0.3)
Gestational diabetes (%)	3 (0.3)
Gestational age at delivery (mean, weeks) (SD)	40.0 (1.2)
Birthweight (mean, grams) (SD)	3562 (477)
Macrosomia (%)	177 (17.3)
Large-for-gestational-age birth (%)	113 (11.1)

Table 2

Predicting large-for-gestational-age (birthweight > 90th percentile) using estimated fetal weight based on 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Ultrasound exams	With covariates		Without covariates	
	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)
T1	0.094 (0.075, 0.109)	0.716 (0.635, 0.756)	0.098 (0.082, 0.115)	0.531 (0.486, 0.581)
T2	0.091 (0.075, 0.107)	0.735 (0.690, 0.795)	0.096 (0.079, 0.112)	0.682 (0.561, 0.741)
T3	0.079 (0.064, 0.092)	0.843 (0.797, 0.880)	0.080 (0.066, 0.094)	0.829 (0.775, 0.873)
T4	0.069 (0.055, 0.084)	0.889 (0.851, 0.923)	0.071 (0.058, 0.087)	0.878 (0.838, 0.914)
T1, T2	0.090 (0.077, 0.104)	0.742 (0.683, 0.797)	0.095 (0.076, 0.108)	0.664 (0.597, 0.731)
T2, T3	0.083 (0.069, 0.096)	0.807 (0.759, 0.850)	0.087 (0.071, 0.102)	0.777 (0.708, 0.825)
T3, T4	0.072 (0.060, 0.087)	0.885 (0.847, 0.918)	0.073 (0.059, 0.089)	0.873 (0.836, 0.910)
T1, T2, T3	0.080 (0.066, 0.094)	0.823 (0.777, 0.867)	0.085 (0.069, 0.097)	0.798 (0.754, 0.843)
T2, T3, T4	0.077 (0.062, 0.089)	0.860 (0.813, 0.887)	0.079 (0.065, 0.092)	0.844 (0.807, 0.888)
T1, T2, T3, T4	0.076 (0.060, 0.088)	0.869 (0.840, 0.906)	0.078 (0.061, 0.094)	0.857 (0.820, 0.901)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Table 3

Predicting macrosomia (birthweight > 4,000 g) using estimated fetal weight based on 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Ultrasound exams	With covariates		Without covariates	
	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)
T1	0.133 (0.116, 0.148)	0.710 (0.659, 0.753)	0.143 (0.120, 0.158)	0.518 (0.475, 0.566)
T2	0.129 (0.111, 0.145)	0.736 (0.695, 0.783)	0.139 (0.118, 0.156)	0.665 (0.597, 0.710)
T3	0.115 (0.100, 0.129)	0.819 (0.783, 0.847)	0.117 (0.103, 0.132)	0.795 (0.758, 0.833)
T4	0.099 (0.083, 0.113)	0.875 (0.846, 0.901)	0.104 (0.085, 0.118)	0.862 (0.824, 0.895)
T1, T2	0.128 (0.109, 0.143)	0.738 (0.690, 0.778)	0.137 (0.118, 0.152)	0.663 (0.617, 0.708)
T2, T3	0.120 (0.101, 0.130)	0.791 (0.754, 0.828)	0.126 (0.110, 0.140)	0.759 (0.704, 0.803)
T3, T4	0.102 (0.084, 0.114)	0.875 (0.846, 0.908)	0.105 (0.089, 0.118)	0.858 (0.822, 0.881)
T1, T2, T3	0.117 (0.098, 0.132)	0.803 (0.761, 0.841)	0.123 (0.105, 0.138)	0.777 (0.729, 0.815)
T2, T3, T4	0.108 (0.094, 0.123)	0.846 (0.804, 0.872)	0.110 (0.095, 0.127)	0.825 (0.784, 0.870)
T1, T2, T3, T4	0.107 (0.092, 0.119)	0.855 (0.825, 0.884)	0.110 (0.094, 0.124)	0.841 (0.800, 0.875)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Table 4

Reclassification based on large-for-gestational-age (birthweight > 90th percentile) using estimated fetal weight based on 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

		T1, T2, T3, T4					
		< 10%		10%–35%		> 35%	
		Median	(95% CI)	Median	(95% CI)	Median	(95% CI)
With covariates							
	< 10%	0.968	(0.933, 0.982)	0.319	(0.194, 0.423)	0.016	(0.000, 0.095)
	10%–35%	0.031	(0.016, 0.056)	0.578	(0.466, 0.746)	0.176	(0.071, 0.333)
Without covariates							
	> 35%	0.000	(0.000, 0.004)	0.090	(0.008, 0.164)	0.789	(0.441, 0.864)
	< 10%	0.957	(0.892, 0.978)	0.319	(0.150, 0.460)	0.028	(0.000, 0.125)
T4	10%–35%	0.043	(0.021, 0.071)	0.573	(0.444, 0.723)	0.182	(0.045, 0.286)
	> 35%	0.000	(0.000, 0.005)	0.103	(0.036, 0.167)	0.789	(0.611, 0.882)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Table 5

Reclassification based on macrosomia (birthweight > 4,000 g) using estimated fetal weight from 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

		T1, T2, T3, T4			
		< 10% Median (95% CI)	10%–35% Median (95% CI)	> 35% Median (95% CI)	
With covariates	< 10%	0.957 (0.834, 0.983)	0.269 (0.100, 0.389)	0.011 (0.000, 0.033)	
	T4 10%–35%	0.040 (0.011, 0.156)	0.641 (0.494, 0.821)	0.174 (0.082, 0.276)	
	> 35%	0.000 (0.000, 0.006)	0.084 (0.050, 0.127)	0.819 (0.684, 0.893)	
Without covariates	< 10%	0.938 (0.764, 0.968)	0.276 (0.071, 0.402)	0.015 (0.000, 0.033)	
	T4 10%–35%	0.058 (0.030, 0.217)	0.633 (0.492, 0.876)	0.194 (0.118, 0.429)	
	> 35%	0.000 (0.000, 0.007)	0.085 (0.020, 0.124)	0.795 (0.549, 0.865)	

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Table A1

Predicting large-for-gestational-age (birthweight > 90th percentile) using biparital diameter, middle abdominal diameter and femur length based on 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Ultrasound exams	With covariates		Without covariates	
	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)
T1	0.093 (0.076, 0.106)	0.715 (0.657, 0.759)	0.098 (0.082, 0.113)	0.564 (0.496, 0.621)
T2	0.089 (0.074, 0.115)	0.757 (0.691, 0.818)	0.095 (0.076, 0.111)	0.679 (0.609, 0.746)
T3	0.078 (0.064, 0.090)	0.859 (0.811, 0.898)	0.079 (0.067, 0.093)	0.842 (0.795, 0.879)
T4	0.069 (0.054, 0.085)	0.894 (0.852, 0.927)	0.071 (0.057, 0.085)	0.882 (0.841, 0.917)
T1, T2	0.091 (0.075, 0.106)	0.759 (0.704, 0.804)	0.095 (0.082, 0.109)	0.674 (0.586, 0.743)
T2, T3	0.084 (0.068, 0.096)	0.836 (0.763, 0.873)	0.084 (0.071, 0.099)	0.819 (0.734, 0.856)
T3, T4	0.070 (0.059, 0.085)	0.893 (0.852, 0.916)	0.073 (0.058, 0.087)	0.878 (0.845, 0.908)
T1, T2, T3	0.081 (0.065, 0.094)	0.835 (0.780, 0.877)	0.084 (0.068, 0.095)	0.815 (0.768, 0.865)
T2, T3, T4	0.074 (0.059, 0.085)	0.882 (0.852, 0.912)	0.075 (0.062, 0.088)	0.871 (0.834, 0.911)
T1, T2, T3, T4	0.072 (0.058, 0.085)	0.885 (0.851, 0.917)	0.073 (0.062, 0.089)	0.873 (0.837, 0.914)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Table A2

Predicting macrosomia (birthweight > 4,000 g) using biparital diameter, middle abdominal diameter and femur length based on 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Ultrasound exams	With covariates		Without covariates	
	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)	Mean-squared error Median (95% CI)	Area under the curve Median (95% CI)
T1	0.131 (0.113, 0.148)	0.711 (0.666, 0.759)	0.143 (0.122, 0.158)	0.549 (0.510, 0.584)
T2	0.127 (0.108, 0.144)	0.753 (0.706, 0.796)	0.136 (0.116, 0.158)	0.677 (0.627, 0.736)
T3	0.110 (0.095, 0.122)	0.844 (0.812, 0.873)	0.113 (0.097, 0.128)	0.820 (0.777, 0.859)
T4	0.098 (0.080, 0.113)	0.882 (0.844, 0.912)	0.102 (0.082, 0.116)	0.866 (0.829, 0.902)
T1, T2	0.128 (0.108, 0.142)	0.752 (0.706, 0.789)	0.137 (0.118, 0.152)	0.671 (0.586, 0.728)
T2, T3	0.116 (0.096, 0.127)	0.824 (0.761, 0.854)	0.119 (0.103, 0.130)	0.804 (0.744, 0.843)
T3, T4	0.098 (0.081, 0.109)	0.884 (0.855, 0.909)	0.103 (0.084, 0.114)	0.867 (0.834, 0.900)
T1, T2, T3	0.112 (0.099, 0.124)	0.829 (0.792, 0.859)	0.118 (0.098, 0.129)	0.802 (0.748, 0.839)
T2, T3, T4	0.099 (0.087, 0.113)	0.874 (0.844, 0.897)	0.103 (0.088, 0.115)	0.860 (0.823, 0.888)
T1, T2, T3, T4	0.102 (0.087, 0.114)	0.873 (0.844, 0.905)	0.105 (0.088, 0.114)	0.860 (0.834, 0.892)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Reclassification based on large-for-gestational-age (birthweight > 90th percentile) using biparital diameter, middle abdominal diameter and femur length from 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Table A3

		T1, T2, T3, T4					
		< 10%		10%–35%		> 35%	
		Median	(95% CI)	Median	(95% CI)	Median	(95% CI)
With covariates	< 10%	0.971	(0.762, 0.996)	0.314	(0.051, 0.448)	0.000	(0.000, 0.065)
	T4 10%–35%	0.027	(0.004, 0.198)	0.612	(0.456, 0.798)	0.217	(0.044, 0.357)
	> 35%	0.000	(0.000, 0.004)	0.066	(0.021, 0.181)	0.772	(0.583, 0.911)
Without covariates	< 10%	0.958	(0.686, 0.992)	0.333	(0.019, 0.436)	0.000	(0.000, 0.071)
	T4 10%–35%	0.038	(0.004, 0.261)	0.604	(0.457, 0.851)	0.216	(0.063, 0.393)
	> 35%	0.000	(0.000, 0.008)	0.061	(0.014, 0.165)	0.765	(0.465, 0.917)

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.

Reclassification based on macrosomia (birthweight > 4,000 g) using biparital diameter, middle abdominal diameter and femur length from 100 bootstrap cross-validations with a 40–60 split between the training and test sets, respectively.

Table A4

		T1, T2, T3, T4			
		< 10% Median (95% CI)	10%–35% Median (95% CI)	> 35% Median (95% CI)	
With covariates	< 10%	0.967 (0.819, 1.000)	0.272 (0.070, 0.414)	0.000 (0.000, 0.022)	
	T4 10%–35%	0.027 (0.000, 0.137)	0.668 (0.508, 0.819)	0.180 (0.111, 0.282)	
	> 35%	0.000 (0.000, 0.006)	0.069 (0.021, 0.150)	0.809 (0.692, 0.886)	
Without covariates	< 10%	0.945 (0.689, 0.993)	0.274 (0.013, 0.396)	0.000 (0.000, 0.027)	
	T4 10%–35%	0.054 (0.007, 0.293)	0.660 (0.482, 0.881)	0.216 (0.103, 0.389)	
	> 35%	0.000 (0.000, 0.006)	0.061 (0.013, 0.142)	0.773 (0.558, 0.862)	

T1, T2, T3, T4 represent ultrasound exams at approximately 17, 25, 33 and 37 weeks of gestation, respectively.

* Covariates include maternal age, prepregnancy body mass index, previous history of small-for-gestational-age birth, smoking during pregnancy.